

formation - QUALIFICATION - emploi

centre d'études  
et de recherches  
sur les qualifications

DOCUMENT N° 18

OXOZJ(XO)

Méthode de comparaison  
des structures d'emploi  
dans un secteur d'activité économique

février 1973

**18**

**Méthode de comparaison  
des structures d'emploi  
dans un secteur d'activité économique**

**février 1973**

S O M M A I R E

AVANT-PROPOS .....	3
<u>CHAPITRE PRELIMINAIRE : L'INFORMATION TRAITEE : PROPOSITION POUR LA CONSTITUTION D'UN FICHER HISTORIQUE</u> .....	7
I - Les principes généraux .....	9
A) Le fichier annuel de référence .....	9
B) Les fiches individuelles d'établissement .....	12
II - Les modalités de réalisation .....	16
A) Les étapes successives .....	16
B) La technique de réalisation .....	19
<u>CHAPITRE I : ANALYSE DES DONNEES : FREQUENCE D'UTILISATION DES RUBRIQUES DE LA NOMENCLATURE</u> .....	29
I - Les problèmes généraux .....	31
II - L'analyse des fréquences d'utilisation .....	36
III - La structure déclarée et la taille des établissements .....	44
<u>CHAPITRE II : COMPARAISON STRUCTURELLE ET ANALYSE QUALITATIVE : RECHERCHE EMPIRIQUE</u> ..	63
I - Représentation structurelle et distance entre emplois .....	65
II - Comparaison structurelle et analyse qualitative .....	84
<u>CHAPITRE III: RECHERCHE THEORIQUE POUR LA CONSTRUCTION D'UN OUTIL DE COMPARAISON</u> .....	101
I - La structure des emplois .....	103
A) Modalités de représentation structurelle .....	105
B) Les justifications économiques de la démarche .....	116
II - Les comparaisons interstructurelles .....	125
A) Distance interstructurelle .....	128
B) Un cadre de recherche explicative .....	138
<u>CHAPITRE IV : MODELE D'EXPLORATION DES DISTORSIONS INTERSTRUCTURELLES : FORMALISATION MATHEMATIQUE DU PROBLEME</u> .....	149
I - Comparaison des structures d'emploi .....	152
A) Exposé théorique des problèmes .....	152
B) Les implantations de la méthode .....	167
II - L'analyse de l'ensemble des structures d'emploi .....	181

AVANT - PROPOS

Les recherches de nature économétrique et statistique poursuivies dans le Département des Synthèses du CEREQ (Centre d'Etudes et de Recherches sur les Qualifications) ont pour objectif dans tous les domaines intéressant l'amélioration des besoins de formation, de tenter de systématiser les approches qualitatives ou quantitatives effectuées soit dans ce même Département, soit dans les autres Départements du CEREQ.

La nouveauté de ces recherches et la nécessité d'intégrer un grand nombre de données ont exigé la mise au point de méthodes d'analyse originales mettant en oeuvre un grand nombre d'investigations statistiques aussi représentatives que possible d'une réalité très complexe.

En effet, la recherche objective des relations entre la formation et l'emploi nécessaires à la définition des besoins de formation est très nouvelle. Elle nécessite à ce titre, en attendant que les critères essentiels de définition de ces relations soient trouvés, l'analyse du plus grand nombre de facteurs susceptibles d'entrer dans ces dernières.

Ces analyses sont complexes, car elles demandent que soient prises en considération simultanément les caractéristiques des emplois, mais aussi de l'environnement de ces emplois et des individus les occupant.

C'est dire que toute recherche dans ce domaine exige que le plus grand nombre de variables soient saisies en même temps et donc que des méthodes d'analyse multicritères soient utilisées.

De ce fait, si une partie des recherches du Département s'est délibérément orientée, dès la création du Centre, vers ce type d'analyse, les méthodes employées ont dû cependant tenir compte de l'information forcément limitée disponible et de la progressivité des analyses causales effectuées par ailleurs.

Ainsi, les relations recherchées dans le présent document ne concernent que les liaisons entre emplois et entre ceux-ci et leur environnement et n'appréhendent qu'une partie des relations formation-emploi, puisqu'en théorie on peut considérer que ces relations sont abordées en distinguant : l'analyse et la description des emplois (décomposition en tâches élémentaires), l'étude des relations entre les tâches et leur environnement (sous l'influence de la technologie ou de l'organisation du travail) enfin la recherche des relations entre les tâches et les individus susceptibles d'occuper les emplois.

En outre, l'utilisation de ces méthodes a été effectuée dans un premier temps à seule fin de mettre en évidence des relations à caractère statistique entre variables, la recherche des relations plus explicatives ne sera abordée que dans un deuxième temps en fonction de l'état d'avancement des recherches correspondantes dans le Centre.

Ces contraintes forcément limitatives ont conditionné la confection de l'outil de réflexion proposé dans ce document qui reste en tous points remarquable.

En effet, si les méthodes d'analyse multicritères peuvent être considérées comme classiques, les auteurs de la présente étude ont fait preuve d'une double originalité :

Ils ont tenté en premier lieu de rendre aussi neutre que possible la méthode d'analyse élaborée à l'égard des liaisons existantes (cette neutralité est de façon générale difficile à obtenir);

- . Ils proposent à propos du problème spécifique des qualifications une notion de distance entre ensembles d'emplois, susceptible d'être améliorée à l'occasion d'études plus qualitatives menées par le CEREQ.

Enfin, il convient de noter que l'exposé présenté revêt uniquement un caractère méthodologique et se rapporte à la recherche des liaisons entre les emplois repérés en termes classiques (dans la nomenclature employée pour l'enquête annuelle sur la structure des emplois du ministère du Travail, de l'Emploi et de la Population) et l'environnement constitué par les établissements ou entreprises industriels.

Les résultats obtenus, quantifiés en fonction des informations statistiques progressivement disponibles, feront l'objet d'une publication ultérieure. En pratique, deux types de résultats seront recherchés :

- . la constitution d'ensembles de variables aussi homogènes que possible pour mettre en évidence les relations essentielles entre les variables considérées comme les plus fondamentales et celles nécessaires à la définition des besoins de formation;

- . des nomenclatures d'emplois plus pertinentes eu égard à l'objectif de recherche des besoins de formation en fonction des ensembles de situations d'emplois retenus comme caractéristiques.

CHAPITRE PRELIMINAIRE

---

L'INFORMATION TRAITEE : PROPOSITION POUR  
LA CONSTITUTION D'UN FICHIER HISTORIQUE

## I - LES PRINCIPES GENERAUX

### A. Le fichier annuel de référence

#### 1) Les objectifs poursuivis

Face aux problèmes relatifs aux prévisions en besoins de main d'oeuvre et en moyens de formation adaptés à l'évolution économique, divers outils ont pu être élaborés, visant, dans le domaine de la collecte de l'information, aussi bien que dans celui de la mise au point des techniques de projection à moyen et à long terme, à éclairer les décisions susceptibles de faciliter l'équilibre qualitatif de l'emploi.

L'enquête sur la structure des emplois, entreprise annuellement depuis le 1er janvier 1968, est destinée à apporter les informations nécessaires à l'étude de la répartition des salariés selon leur formation, leur qualification et leur spécialité exercée. Les quatre enquêtes actuellement disponibles (1968, 1969, 1970 et 1971) permettent de suivre l'évolution de cette répartition dans le temps et de tracer une première esquisse globale des grandes tendances qui se dessinent, point de départ d'une analyse plus approfondie ayant pour vocation la prévision de l'emploi par profession dans chaque secteur d'activité.

Deux orientations peuvent être envisagées :

- Améliorer d'abord les techniques de projections globales, mais les efforts développés dans cette direction ne semblent pas devoir déboucher sur un gain d'efficacité majeur.

- Ou réorienter l'analyse autour d'une idée nouvelle, selon laquelle l'évolution et la transformation d'un secteur s'expliquent nécessairement par l'évolution et la transformation des entreprises qui le composent. C'est dans ce sens que nous avons engagé cette étude, en nous limitant uniquement à la mécanique.



## 2) Le fichier annuel de référence

Les données de base, nécessaires à l'accomplissement de nos travaux, n'existent pas à l'heure actuelle sous la forme souhaitée, mais il semble parfaitement possible de les constituer à partir de l'information disponible.

Chaque année, en effet, les établissements occupant plus de 10 salariés, sont tenus de répondre à l'enquête sur la structure des emplois : la population ainsi contactée se trouve, sauf exceptions liées principalement aux créations et aux disparitions, composée des mêmes établissements, dont il est possible de retracer l'évolution et la transformation dans le temps.

La restructuration de l'information brute permet dès lors de construire un fichier annuel des établissements de la mécanique au cours de la période 1968-1971, indiquant la répartition de leur personnel salarié entre les rubriques de la nomenclature utilisée.

Ces établissements appartiennent à des entreprises qui, selon le cas, regroupent une ou plusieurs unités productives. Si l'on connaît tous les établissements dépendant de chaque entreprise, on est en mesure de construire, par regroupement des données individuelles, un fichier annuel des établissements. L'information devant permettre le passage de l'établissement à l'entreprise existe (notamment par l'Enquête Annuelle d'Entreprises du Ministère du Développement Industriel et Scientifique) et nous nous proposons de l'utiliser.

De façon plus précise, l'établissement apparaîtra toujours comme l'unité de base fondamentale, mais, dans tous les cas et chaque année, son entreprise d'appartenance sera mentionnée. Cela revient à mettre au point, dans une première étape, un fichier annuel de référence, composé d'établissements dont le regroupement permettra, dans une deuxième étape de créer un fichier d'entreprises.

3) Le champ couvert (critère d'activité)

Il convient au préalable de lever une difficulté tenant à la définition du champ de l'étude. Les notions de secteur d'entreprise et de secteur d'établissement ne se confondent pas nécessairement, certains établissements de la mécanique pouvant appartenir à des entreprises des secteurs hors mécanique, et certaines entreprises de la mécanique pouvant posséder des établissements des secteurs hors mécanique.

Les objectifs poursuivis devraient normalement nous conduire à négliger tous les établissements appartenant à des entreprises des secteurs hors mécanique, le critère d'entreprise étant prépondérant. Il nous a semblé cependant souhaitable, pour des raisons tenant principalement aux possibilités d'utilisation ultérieure de l'information, de les inclure dans la population traitée.

Le fichier annuel de référence sera dès lors composé de l'ensemble des établissements appartenant :

- au fichier annuel d'entreprises de la mécanique;
- au fichier annuel d'établissements de la mécanique.

Il regroupera donc :

a) l'ensemble des établissements de la mécanique se confondant avec une entreprise de la mécanique (entreprises à établissement unique);

b) l'ensemble des établissements de la mécanique dépendant d'une entreprise de la mécanique (entreprises à établissements multiples);

c) l'ensemble des établissements de la mécanique dépendant d'une entreprise hors mécanique;

d) l'ensemble des établissements hors mécanique dépendant d'une entreprise de la mécanique.

Le regroupement des sous-populations a,b,c, définit le secteur d'établissement de la mécanique. Le regroupement des sous-populations a, b et d, définit le secteur d'entreprise de la mécanique.

*B - Les fiches individuelles d'établissement*

1) Le fichier induit

L'enquête sur la structure des emplois, effectuée par le Ministère du Travail, de l'Emploi et de la Population et par l'INSEE, fournit les données relatives à la répartition de la population active par type et niveau de qualification professionnelle pour tous les établissements entrant dans son champ d'application. Le critère de l'activité économique exercée, immédiatement repérable grâce au numéro INSEE, permet d'opérer une première classification des établissements et d'isoler notamment ceux du secteur mécanique.

L'enquête annuelle d'entreprises, effectuée par le Ministère du Développement Industriel et Scientifique, fournit les données relatives à la structure des entreprises par établissements. La connaissance des numéros INSEE permet d'opérer une classification des établissements et des entreprises selon leur appartenance aux secteurs mécanique et hors mécanique. Le fichier des entreprises INSEE apporte, pour une seule année, un renseignement similaire, complété toutefois par l'identification en clair (nom et adresse des établissements et des entreprises).

L'exploitation du fichier d'entreprises du secteur mécanique peut conduire à la génération d'un fichier induit d'établissements appartenant à des entreprises de la mécanique, et, selon les cas, aux secteurs mécanique ou hors mécanique.

Il est par ailleurs possible d'extraire du fichier d'entreprise hors mécanique un fichier induit d'établissements de la mécanique, qui, joints aux établissements de la mécanique appartenant à des entreprises de la mécanique, constituent une population directement comparable à celle des établissements de la mécanique issus de l'enquête sur la structure des emplois.

En ce qui concerne les établissements hors mécanique appartenant à des entreprises de la mécanique, le rapprochement des données exige la définition d'une procédure indirecte, visant à rechercher dans le fichier ESE les établissements du champ grâce au fichier EAE.

## 2) L'identificateur d'établissement

Le dénominateur commun entre le fichier induit et le fichier ESE est constitué par le numéro INSEE, qui, d'une façon générale, doit être identique pour un établissement, qu'elle que soit la source d'information utilisée. Tout établissement est dès lors caractérisé en premier lieu par son identificateur, dont la connaissance doit permettre d'extraire des divers fichiers en présence, les données utiles à l'exploitation ultérieure.

Cet élément constitue, au moins dans une première étape, la clé de voûte de toute la procédure de fusion. La structure de l'information brute se présente sous la forme de huit fichiers différents (ESE : 1968, 1969, 1970, 1971 ; EAE : 1968, 1969, 1970 ; INSEE : 1970) ayant pour vocation le recensement des établissements entrant dans leur champ d'application. En négligeant divers problèmes sur lesquels nous reviendrons plus loin, tout identificateur apparaissant au moins une fois dans un des fichiers doit se rapporter à un établissement ayant existé au cours de la période 1968-1971. Ou, si l'on combine les données successives, la liste de tous les identificateurs différents préfigure la population couverte par le fichier annuel de référence.

La démarche n'évite pas absolument la critique : certains établissements peuvent très bien échapper à toute investigation et surtout les identificateurs erronés risquent de conduire à des recherches portant sur des établissements purement fictifs. En ce qui concerne le premier point, nous pensons que la multiplication des sources permettra une approche satisfaisante de l'exhaustivité.

Quant aux erreurs d'identification, elles devront être détectées par une analyse des résultats et corrigées par regroupement des informations sous un identificateur arbitraire indépendant des caractéristiques retenues dans le numéro INSEE.

### 3) Les fiches individuelles

Dans le cas de la constitution d'un fichier "à la main", une fiche est créée pour tous les établissements rencontrés dans l'exploitation des sources d'information retenues, et les renseignements disponibles sont portés dans les cases correspondantes de la fiche qu'il s'agit de remplir.

La démarche utilisée ici sera similaire. Lorsqu'elles sont disponibles, les informations recherchées seront rangées dans un ordre déterminé derrière chacun des identificateurs repérés. La structure globale de l'information se retrouve donc au niveau individuel : pour un établissement quelconque, on obtient en effet dans le meilleur des cas, huit grands types de données émanant des divers fichiers utilisés. Il devient dès lors parfaitement possible de sortir les résultats sous forme de fiches individuelles réunissant les renseignements propres à chaque unité productive existante.

Pour faciliter l'analyse des premiers résultats, nous nous limiterons au départ à une information minimale, de façon à simplifier la procédure sans nuire à l'efficacité des recherches. Il suffit en effet de connaître principalement le nom et l'adresse de l'établissement (INSEE), le numéro de l'entreprise à laquelle l'établissement est rattaché chaque année (EAE, INSEE), son effectif annuel global fourni par les diverses sources et la structure annuelle condensée de son personnel salarié (ESE).

En classant les "fiches" en fonction de la localisation des établissements (régions, départements, communes) et selon les modalités de leur remplissage (mise en évidence des données com-

plémentaires notamment), on peut penser, sinon pouvoir résoudre tous les problèmes existants, au moins organiser le support d'une enquête supplémentaire auprès des Echelons Régionaux, visant à améliorer les informations disponibles.

## II - LES MODALITES DE REALISATION

### A. Les étapes successives

#### 1) L'intersection des champs

L'enquête annuelle d'entreprises, effectuée auprès des entreprises dont l'activité principale appartient exclusivement aux secteurs industriels, ne permet pas d'isoler les établissements de la mécanique dépendant d'une entreprise à activité principale non industrielle. De tels établissements, s'il en existe et s'ils entrent dans le champ d'application de l'enquête sur la structure des emplois ne pourront être que partiellement renseignés (ESE, INSEE)

Par ailleurs, les modalités de réalisation de l'enquête annuelle d'entreprises diffèrent selon la taille des unités questionnées. Lorsqu'une entreprise occupe 50 salariés et plus, elle est tenue de remplir et de retourner un questionnaire général (QG) se rapportant d'une part à l'entreprise dans son ensemble et d'autre part à chacun des établissements qu'elle possède. Cet éclatement des données, qui permet la génération du fichier d'établissements induit, n'existe pas pour les petites entreprises. Le questionnaire simplifié (QS) adressé à toutes les entreprises de 20 à 49 salariés et à une fraction des entreprises de 6 à 19 salariés, n'interroge pas en effet sur les établissements, mais uniquement sur l'entreprise : l'information disponible n'est pas, en ce qui les concerne, directement utilisable.

Ainsi, sous réserve d'appartenance de l'entreprise au secteur industriel, tous les établissements de 50 salariés et plus entrent nécessairement dans le champ d'application des enquêtes annuelles d'entreprises et sur la structure des emplois. Le problème est plus complexe pour les établissements de taille inférieure. Ils doivent

appartenir à une entreprise de 50 salariés et plus (critère de taille EAE) et occuper plus de 10 salariés (critère de taille ESE). A défaut, ils échappent selon les cas à l'une des deux sources (établissement de moins de 10 salariés appartenant à une entreprise de plus de 50 salariés; établissement de plus de 10 salariés appartenant à une entreprise de moins de 50 salariés) ou aux deux sources réunies (établissement de moins de 10 salariés appartenant à une entreprise de moins de 50 salariés).

## 2 - La procédure complémentaire

Le passage à l'entreprise, par regroupement des données individuelles d'établissements, peut parfaitement être effectué pour les entreprises de plus de 50 salariés possédant exclusivement des établissements de plus de 10 salariés.

Il n'en va pas de même pour :

- toutes les entreprises possédant un ou plusieurs établissements de moins de 10 salariés, même si la taille de l'entreprise est supérieure ou égale à 50 (champ EAE);

- toutes les entreprises de moins de 50 salariés, même si chacun de leurs établissements occupe plus de 10 salariés (champ ESE).

L'application stricte des deux critères de taille (établissement, entreprise) risque donc de conduire à éliminer en bloc toutes les entreprises pour lesquelles une fraction de l'information recherchée sort du domaine d'intersection du champ des deux enquêtes. La sélection peut paraître dans certains cas extrêmement sévère. Aussi sommes-nous tentés de définir une procédure complémentaire visant à se procurer l'information manquante lorsqu'elle présente un caractère marginal par rapport à l'information connue.

Deux démarches peuvent être envisagées :



- procéder à une ventilation par établissements des données relatives aux entreprises de moins de 50 salariés (recours au fichier INSEE);

- procéder à une ventilation par postes d'emploi du personnel des petits établissements (moins de 10 salariés) appartenant à des entreprises moyennes ou grandes (plus de 50 salariés).

La première solution a pour objet d'assurer une plus grande homogénéité des données par rapport au critère établissement. La deuxième solution a pour objet d'assurer une plus grande homogénéité des données par rapport au critère entreprise.

### 3) Les étapes de la réalisation

L'adoption d'une telle procédure, en l'absence de toute contrainte relative à la délimitation du champ de l'étude, paraît susceptible de faire progresser, de proche en proche, la collecte de l'information vers l'exhaustivité. En pratique, les étapes de cette progression s'accompagnent d'une lourdeur croissante d'exécution et d'un enchaînement cumulatif des difficultés de réalisation.

Aussi, dans un premier temps, semble-t-il préférable de nous limiter à la création d'un fichier annuel de référence, composé uniquement des établissements appartenant :

- au fichier annuel d'entreprises de la mécanique occupant 50 salariés et plus (sources : EAE, INSEE);

- au fichier annuel d'établissements de la mécanique occupant 10 salariés et plus (sources : ESE, INSEE).

Dans un deuxième temps, la fusion des informations pourra être étendue à l'ensemble des établissements appartenant aux entreprises occupant 20 salariés et plus; et éventuellement, dans un

troisième temps, aux entreprises occupant plus de 10 salariés.

D'une façon générale, chaque fois qu'une entreprise, entrant dans le champ d'application, possède un établissement de moins de 10 salariés, il apparaît nécessaire de procéder à la ventilation du personnel de cet établissement par postes d'emplois. On peut estimer que ce cas est d'autant plus rare que la taille de l'entreprise est grande, et qu'il sera peu fréquent dans la première phase de réalisation de l'étude. En revanche, la recherche d'informations complémentaires peut présenter, dans la deuxième et surtout dans la troisième phase de l'étude, un caractère prohibitif, excluant la poursuite des travaux dans cette direction : nous nous limiterons donc probablement à ce stade aux seules informations disponibles.

#### B - La technique de réalisation

##### 1) Les informations fusionnées

La démarche utilisée dans la première étape est résumée par l'organigramme I. Elle consiste, après transformation des données brutes disponibles (génération du fichier induit des établissements), à réunir sous un même numéro d'identification toutes les informations possédées sur les établissements entrant dans le champ de l'étude. Les établissements de la mécanique sont repérés par les trois sources : EAE (31.XII.1968 ; 31.XII 1969 ; 31.XII 1970), INSEE (1970), ESE (1.I.1968 ; 1.I.1969 ; 1.III.1970, 1.III.1971). Les établissements hors mécanique, dépendant d'une entreprise de la mécanique, ne peuvent être repérés que par les sources EAE et INSEE, et il s'agira, une fois définie la liste des identificateurs, d'extraire les informations correspondantes dans ESE. On obtient bien ainsi un fichier annuel de référence composé : d'une part des établissements appartenant aux entreprises de la mécanique occupant 50 salariés et plus ; d'autre part les

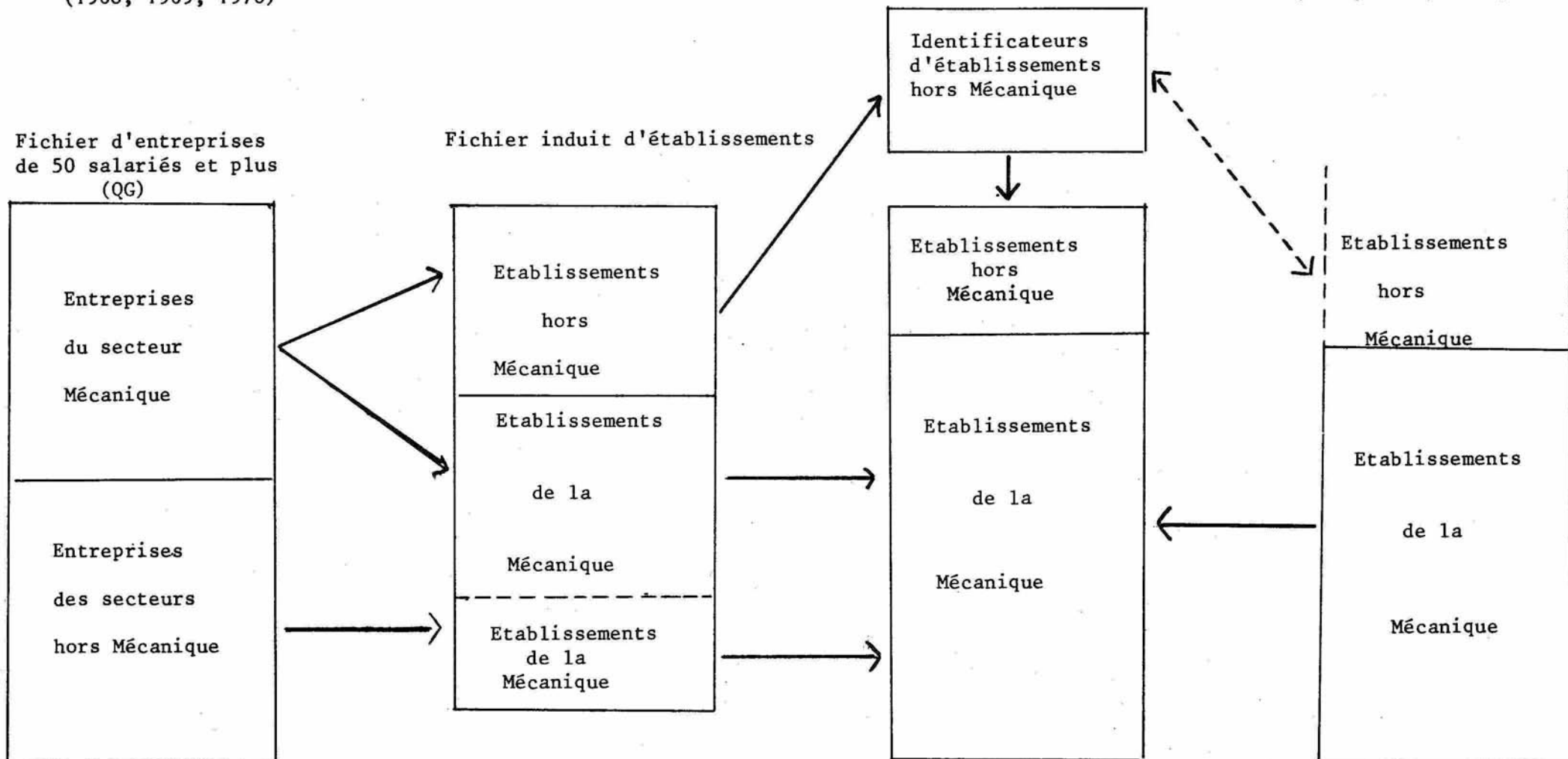
ORGANIGRAMME I

E.A.E.

(1968, 1969, 1970)

E.S.E.

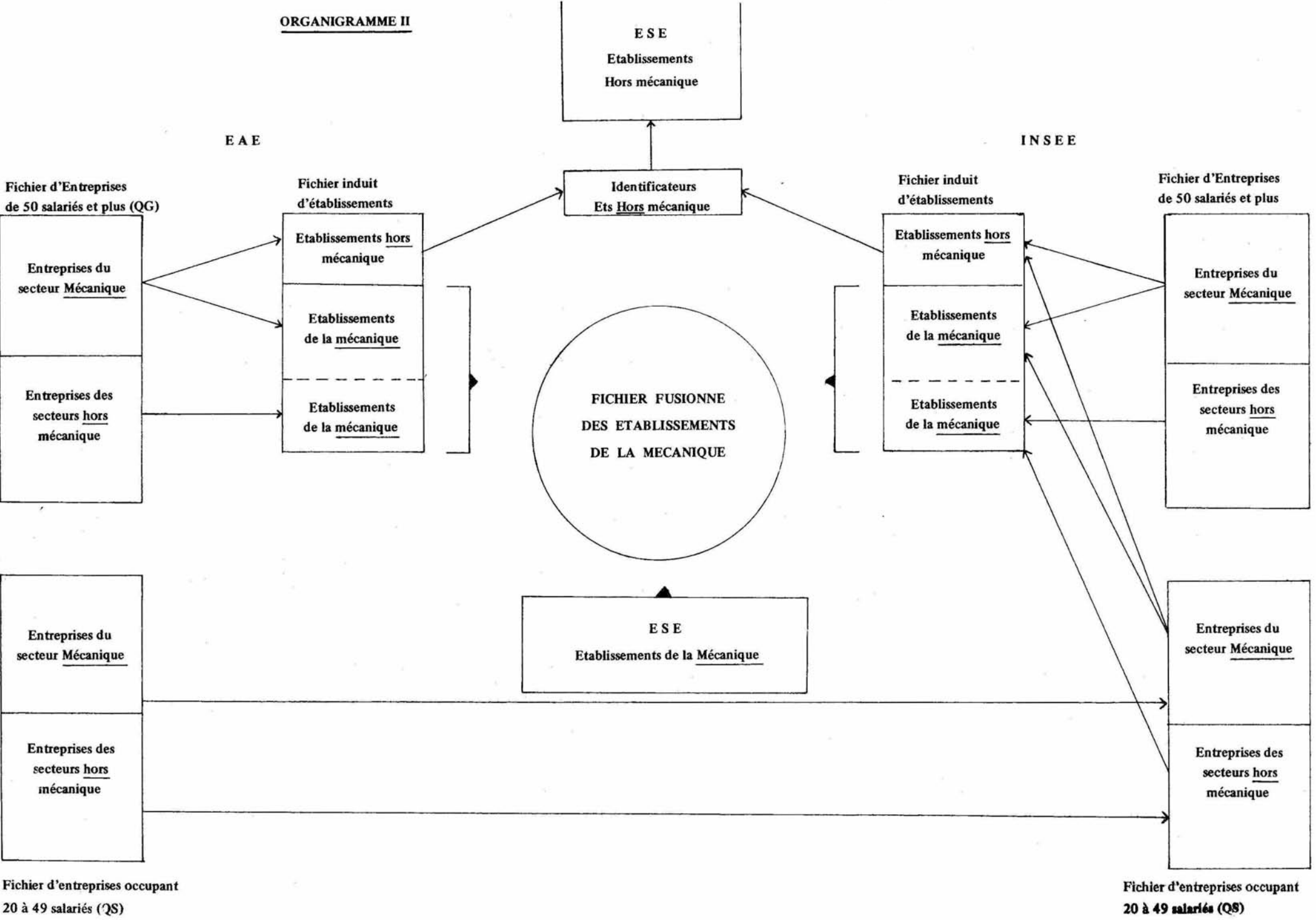
(1968, 1969, 1970, 1971)



Fichier d'entreprises de 50 salariés et plus

Fichier induit d'établissements

**ORGANIGRAMME II**



établissements de la mécanique occupant 10 salariés et plus.

Les opérations à effectuer dans le cadre de la deuxième phase de l'étude sont décrites par l'organigramme II, dont une partie (relative aux entreprises de plus de 50 salariés ) coïncide avec l'organigramme I, les autres opérations étant spécifiques à l'extension du champ d'application aux entreprises occupant 20 à 49 salariés.

L'utilisation des données EAE non ventilées par établissements (QS) exige leur transformation préalable : la structure des entreprises par établissements, fournie par le fichier INSEE, donne une grille de passage relative à une seule année. Nous serons donc amenés à supposer que toutes les entreprises existant en 1968, 1969 et 1970 (source EAE) possèdent au cours de ces trois années les mêmes établissements qu'en 1970 (source INSEE). Ce n'est qu'au prix d'une telle hypothèse qu'il sera possible de constituer un fichier annuel de référence étendu aux entreprises de plus de 20 salariés.

## 2) Le fichier de travail

L'information brute de départ se présente sous la forme de huit fichiers, différant selon la source d'origine (EAE, INSEE, ESE) et l'année de collecte des données (1968, 1969, 1970, 1971). Il semble dès lors naturel de reproduire les enregistrements des fiches d'origine relatives aux établissements entrant dans le champ d'application de l'étude, sur le modèle suivant :

N° établissement INSEE	Origine de l'information	Année de référence	Données spécifiques de la source
------------------------	--------------------------	--------------------	----------------------------------

Sans fixer définitivement la liste des données spécifiques, le tableau ci-dessous fournit une indication de leur contenu éventuel :

Origine de l'information niveaux	E A E (1968, 1969, 1970)	INSEE (1970)	ESE (1968, 1969, 1970, 1971)
Etablissement	Effectif total	-nom et adresse - lien - tranche d'eff.	effectif total
Entreprise	-N° Identification - effectif total - nombre d'éts.	-N° identificat. -nom et adresse	
Structure des emplois			Répartition des salariés par catégories d'emplois regroupées

Il suffit alors de trier ces informations en fonction du code établissement et de ranger dans un ordre déterminé derrière chacun des identificateurs, les diverses données qui lui sont rattachées.

Cette opération permet d'obtenir :

- un fichier de travail édité par fiches successives selon un modèle pré-établi et destiné à faciliter l'analyse des premiers résultats et la mise au point définitive de l'outil élaboré;

- un fichier de travail enregistré sur bande magnétique et appelé à recevoir les éléments de la mise au point définitive du fichier annuel de référence.

### 3) Vers un fichier définitif

Il n'est pas possible de préjuger à l'avance de la qualité des premiers résultats obtenus. On observera néanmoins que les informations réunies sont, d'une façon générale, uniquement complémentaires et que, de ce fait, les lacunes ou les erreurs provenant de l'une des sources ne pourront en aucune façon être directement comblées ou corrigées par la connaissance des autres sources.

Deux démarches analytiques peuvent être tentées. La comparaison entre elles des fiches partiellement remplies devrait d'abord permettre la mise en évidence des renseignements complémentaires susceptibles d'appartenir à un seul et même établissement : la présomption d'erreur sur l'identificateur sera d'autant plus grande que les données obtenues par ailleurs correspondront entre elles. Ce cas, lié à la modification de l'une des caractéristiques principales de l'établissement (changement d'adresse, d'activité principale notamment), ou plus simplement à une erreur de déclaration ou d'enregistrement, risque d'avoir une fréquence non négligeable. Lorsqu'il sera décelé, on procédera à la fusion des données correspondantes. Pour cela, l'identificateur INSEE sera complété sur le fichier de travail par un identificateur CEREQ arbitraire, appelé à devenir prépondérant sur le fichier définitif : dans le cas de la fusion, l'un des identificateurs arbitraires sera effacé et remplacé par l'autre.

Par ailleurs, la complémentarité des sources, pour une même année et dans le temps devrait permettre, sinon de connaître de façon certaine, au moins d'apprécier dans une certaine mesure la période d'existence de l'établissement entre 1968 et 1971 et d'attribuer l'absence d'information à une non-réponse, à un processus de création non encore amorcé, ou à une disparition.

S'il s'agit d'une non-réponse, les données antérieures et postérieures de même type, combinées aux autres données complémentaires, risquent de fournir des éléments utiles à une extrapolation ou une interpolation éventuelle. S'il s'agit d'une création ou d'une disparition, le phénomène sera précisé, codifié, et viendra améliorer la constitution du fichier définitif.



### CONCLUSION

Un des objectifs fondamentaux de l'analyse est la mise en évidence des problèmes liés à la discontinuité de l'information dans le temps. Les résultats qu'il est possible de dégager des données brutes actuelles de l'Enquête sur la structure des Emplois, dépendent du nombre d'établissements contactés, du taux de réponse de ces établissements et de la modification de ces caractéristiques dans le temps. Dans sa conception la plus simple, le passage à la série temporelle peut être obtenu par alignement des données sur les informations les plus pauvres. Une telle démarche consiste à ne retenir dans les études, que les établissements ayant répondu chaque année à l'enquête, c'est-à-dire à ne travailler que sur le champ constant défini comme le sous-ensemble commun aux quatre enquêtes successives.

Satisfaisante dans son principe, cette solution minimale risque, dans son application, de conduire à une perte importante de l'information existante, le processus de sélection éliminant les établissements n'ayant pas répondu au moins une fois à l'enquête; les établissements créés ou disparus au cours de la période d'observation; les établissements ayant changé de numéro d'identification INSEE, reflétant ainsi une modification de leurs caractéristiques ou une erreur de déclaration ou d'enregistrement.

De telles restrictions, qui réduisent fortement le champ d'application des études, aussi bien que la portée des résultats qui s'en dégagent, doivent être levées dans la mesure du possible. A ce titre, le repérage des créations et des disparitions, qui constituent l'un des facteurs essentiels du mouvement de croissance et de transformation structurelle, est susceptible de fournir un élément appréciable d'amélioration de la qualité des informations traitées. L'utilisation du critère de la continuité dans le temps,

inspiré des principes de base de la constitution du fichier des grands Etablissements de l'INSEE, conduit également à étendre le champ d'application des études dynamiques, généralement tributaires de la correspondance entre l'identificateur et les caractéristiques propres de l'établissement.

En ce qui concerne le problème des fiches incomplètes, liées à l'absence de réponse de l'établissement à l'une des enquêtes annuelles, la solution à adopter n'a pas été arrêtée définitivement. Il semblerait souhaitable de pouvoir "comblé le trou" lorsque la fiche est par ailleurs correctement remplie. L'analyse des premiers résultats devrait permettre de se prononcer sur la nature et l'importance des informations manquantes qu'il serait essentiel de posséder. Elle devrait également faciliter la définition des opérations à effectuer pour se les procurer : opérations mécaniques (interpolation, extrapolation), recours à l'enquête complémentaire. Il est d'ores et déjà possible cependant d'avancer qu'un tel travail exige une compétence et une connaissance auxquelles nous ne saurions prétendre : il sera donc demandé aux Echelons Régionaux une étroite collaboration sur les problèmes relevant de leur propre région.

Ce n'est qu'à l'issue de cette phase complexe que le passage au fichier des entreprises sera effectué. Le fichier annuel de référence est conçu pour satisfaire à cet objectif, et les opérations nécessaires à la transformation des données ne devraient pas poser des problèmes majeurs.

*C H A P I T R E I*

---

*ANALYSE DES DONNEES : FREQUENCE D'UTILISATION DES  
RUBRIQUES DE LA NOMENCLATURE*

## I - LES PROBLEMES GENERAUX

### 1. Le principe du regroupement

La nomenclature des emplois, à laquelle se réfèrent les déclarations annuelles des employeurs, est destinée principalement à mettre en évidence les besoins de main d'oeuvre, et par suite de formation, attachés à l'évolution économique et au progrès technique. Il s'agit de mieux connaître l'évolution passée des catégories d'emploi afin de mieux aligner les efforts de formation sur l'appréciation des besoins futurs.

Le personnel des établissements enquêtés doit être réparti selon la grille fournie par les 296 rubriques existantes : ainsi, la structure des emplois obtenue ne peut que refléter la conception ayant présidé à l'élaboration de la nomenclature, indépendamment des critères spécifiques pouvant éventuellement être utilisés pour classer les catégories d'emploi à l'intérieur même de l'établissement et tenant compte notamment du statut professionnel, du salaire, des aptitudes et des qualités particulières du titulaire de chacun des postes.

L'avantage d'une telle procédure est de faciliter les comparaisons individuelles et d'adapter spontanément l'analyse aux objectifs généraux poursuivis. Mais les distorsions susceptibles d'exister entre le modèle préétabli et l'organisation effective de l'unité productive, fortement marquée par l'influence du titulaire sur le contenu de l'emploi lui-même, rendent souvent malaisé l'établissement des déclarations et risquent de déformer dans une certaine mesure le contenu réel de la notion de structure.

Globalement, on peut supposer que les erreurs se compensent ou utiliser des coefficients de redressement propres à chacun des emplois déclarés. Au niveau individuel, la démarche la plus souvent adoptée consiste à regrouper les rubriques entre lesquelles les confusions sont les plus fréquentes.

## 2) Les sources de confusion

Nous voulons adapter le cadre structurel fourni par les déclarations aux objectifs spécifiques que nous poursuivons, c'est-à-dire construire un outil permettant d'analyser en terme de population d'établissements ou d'entreprises la répartition des emplois et sa transformation dans le temps.

Les rubriques correspondent théoriquement, pour une spécialité exercée donnée, à des emplois de même nature ou similaires requérant un certain niveau de qualification et de formation. Les emplois réels présentent des caractéristiques multiples qui leur sont propres ; le déclarant doit isoler celles retenues dans la nomenclature et négliger toutes les autres. En pratique, une telle démarche est délicate et la structure qui est fournie risque de s'écarter plus ou moins sensiblement de la structure qui aurait dû ou qui aurait pu être présentée.

Les confusions s'expriment à trois niveaux différents : un type d'emploi défini par ses caractéristiques propres peut correspondre à plusieurs rubriques (confusion liée à la pluralité des tâches) ou ne pas correspondre à l'une des rubriques existantes (confusion liée à la spécificité des tâches) ; des emplois différents peuvent satisfaire à une seule et même rubrique (confusion liée au choix des critères de similitude). Il s'agit donc de trouver un dénominateur commun entre deux types de découpage, l'un théorique, conforme au contenu des rubriques de la nomenclature, l'autre observé, conforme au contenu des postes de déclarations, sans trop perdre ni déformer l'information disponible.

Les données regroupées par la définition des rubriques ne peuvent être ventilées et la classification des emplois spécifiques doit s'opérer par assimilation aux emplois énumérés : la perte d'information qui en résulte est incompressible, la nomenclature de référence

intervenant comme une contrainte impossible à lever. En ce qui concerne les confusions liées à la pluralité des tâches pour un même emploi, seule l'agrégation des rubriques correspondantes semblerait permettre d'obtenir des résultats significativement comparables. L'équilibre recherché se situe donc largement en deça du niveau de décontraction de la nomenclature.

### 3) Les critères de regroupement

Le principe du regroupement remet en cause la définition du contenu de chaque rubrique en fonction du contenu réel des emplois existant dans les établissements. La solution retenue ne peut être adaptée à tous les cas rencontrés : trop détaillée pour certains, elle ne supprimera qu'imparfaitement les sources de confusion ; trop synthétique pour d'autres, elle conduira à la perte d'une information qui aurait très bien pu être utilisée. Doit-on regrouper les fonctions directives, administratives, commerciales et techniques parce qu'elles peuvent être placées sous la responsabilité d'une seule et même personne dans une entreprise de petite dimension ? On serait tenté d'un autre côté d'utiliser une décontraction plus fine que celle des 296 rubriques pour appréhender la structure d'une grande entreprise.

Les seuls critères objectifs disponibles sont fournis par la fréquence d'utilisation de chacune des rubriques et par le poids que représente en moyenne leur effectif. Le problème se trouve dès lors légèrement déplacé dans la mesure où les critères utilisés ne reflètent pas nécessairement toutes les confusions possibles, alors qu'ils mettent en évidence un élément nouveau, celui des rubriques jamais ou rarement retenues dans les déclarations. On peut choisir de regrouper les rubriques dont la fréquence d'utilisation se situe en-dessous d'un certain seuil, fixé arbitrairement.

Cela ne donne pas pour autant les modalités du regroupement à effectuer. Il faut se référer aux analyses du contenu des déclara-

rations, qui indiquent les erreurs de classement le plus souvent commises et fournissent par là-même une clé de regroupement permettant de les absorber. Il faut définir une grille hiérarchisée des emplois, directement dérivée du découpage apporté par la nomenclature, et mettant en évidence les proximités de fait et les enchaînements possibles des divers emplois entre eux.

#### 4) L'équilibre fonctionnel

La combinaison des critères quantitatifs (mesure des fréquences d'utilisation, évaluation du poids de chaque unité structurelle) et des critères qualitatifs (proximité de fait du contenu des rubriques, appréciation des enchaînements possibles des catégories d'emploi entre elles), doit conduire au tracé d'un premier découpage structural, condensant l'information des rubriques de la nomenclature en un nombre réduit de postes plus significatifs et plus facilement comparables.

Une telle démarche tend à favoriser en pratique la ventilation de certaines catégories d'emplois (ouvriers qualifiés, employés de bureau) au détriment des emplois sans qualification particulière (ouvriers spécialisés, manoeuvres, employés non qualifiés) et, dans une moindre mesure, des emplois exigeant un niveau de formation relativement élevé (cadres supérieurs, cadres techniques et techniciens).

Cela apparaît comme une conséquence logique des principes de la nomenclature : l'objectif poursuivi, mettre en évidence les besoins en formation, n'exige pas une analyse détaillée des emplois "dont la pratique s'acquiert généralement en quelques semaines (ouvriers spécialisés, manoeuvres, employés non qualifiés). En ce qui concerne les autres catégories d'emploi, l'accent est mis principalement sur le niveau de formation requis, qui constitue un indicateur du niveau de qualification, et sur la spécialité exercée, sans que l'on puisse par recouplement isoler de façon satisfaisante les grandes fonctions, dont la répartition, vraisemblablement sous-jacente aux déclarations, servirait mieux les besoins de l'analyse des transformations des entreprises.

Une approche peut cependant être tentée dans cette direction, en s'efforçant de coiffer, pour autant que le découpage en rubrique le permet, les niveaux de qualification et la spécialité exercée, par une classification des emplois en grandes catégories fonctionnelles. Cela revient à jouer à la fois sur la notion de fonction, lorsqu'elle est saisissable, le niveau de qualification, lorsqu'il apparaît, et la spécialité exercée, lorsqu'elle s'avère significative, ces divers critères intervenant de façon inégale et imparfaite, à l'intérieur des limites fixées par la définition des 296 rubriques et par les conditions d'utilisation de la nomenclature par les entreprises.



## II - L'ANALYSE DES FREQUENCES D'UTILISATION

### 1) Les résultats bruts

Bien que, dans leur ensemble, les établissements relevant de l'étude exercent une activité similaire ou voisine (secteur mécanique), le nombre des rubriques utilisées au moins une fois entre 1968 et 1971 par l'un quelconque des établissements enquêtés, est extrêmement élevé : 201 rubriques au total sur les 296 possibles. En sens inverse, aucune des rubriques de la nomenclature des emplois n'est utilisée systématiquement par tous les établissements ; les ouvriers professionnels sur machines (387) et les ouvriers spécialisés (471) dont l'effectif représente pourtant en moyenne respectivement 13 et 16 % du total des emplois, apparaissent dans moins de trois déclarations sur quatre.

D'une façon générale, cinq rubriques seulement (agents de maîtrise d'ouvriers qualifiés du travail des métaux et de l'électricité, ouvriers professionnels sur machine, ouvriers spécialisés, manoeuvres, personnel de secrétariat) sont utilisées par plus de la moitié des établissements, 51 par plus d'un établissement sur dix, 66 par plus d'un établissement sur vingt. A l'opposé, 95 rubriques ne sont jamais utilisées, 184 ne le sont jamais ou rarement et, en définitive, trois rubriques sur quatre sont exclues, de toutes les déclarations ou de la plupart d'entre elles.

La dispersion du choix des rubriques, constatée en première analyse, est donc plus apparente que réelle. Deux résultats semblent se dessiner clairement :

- il n'existe pas de cadre structurel simplifié, obtenu uniquement par sélection des rubriques et tel que toutes les rubriques retenues correspondraient à des emplois existant dans tous les établissements étudiés ;

CLASSIFICATION DES RUBRIQUES SELON LEUR FREQUENCE D'UTILISATION

Rubrique utilisée par	Code E	NB rubriques
+ de 75 % des ét.		0
50 à 75 % des ét.	271;387;471;472;610	5
30 à 49 % des ét.	100, 284, 361, 380, 382 501, 502, 600, 606	9
20 à 29 % des ét.	226; 260; 272; 280; 382 384; 500; 514; 602; 607 611; 612	12
15 à 19 % des ét.	203; 242; 250; 341; 370 388; 391; 601; 604; 608 701; 704	12
10 à 14 % des ét.	275; 276; 288; 360; 362 371; 394; 513; 609; 700 703; 951; 952	13
5 à 9 % des ét.	206; 223; 261; 277;320 340; 372; 389; 463; 470 503; 603; 605; 702;707	15
1 à 4 % des ét.	Liste non détaillée	46
0,01 à 0,09 % ét.	Liste non détaillée	89
Aucun établissement	Liste non détaillée	95

- la nomenclature des emplois à laquelle se réfèrent les déclarations présente une gamme de postes spécifiques sortant d'une façon générale du mode organisationnel des unités productives étudiées.

2) L'emploi exercé et l'activité économique de l'établissement

A l'exception des rubriques du sous-groupe 11 (chef d'une entreprise inscrite au Répertoire des Métiers) et du sous-groupe 21 (cadres techniques supérieurs titulaires d'une licence ou d'un diplôme universitaire supérieur à la licence), les premières ne devant en principe jamais apparaître dans les déclarations (le Répertoire des Métiers ne concerne que les entreprises de moins de 10 salariés), les deuxièmes ayant été très peu retenues (erreur d'affectation par assimilation à une absence de diplôme ou à un diplôme d'ingénieur), les rubriques jamais ou rarement utilisées correspondent d'une façon générale à des spécialités qu'il semble naturel de ne pas trouver associées à l'activité économique des établissements de la mécanique.

Sont ainsi exclues des déclarations par au moins 99 établissements sur 100 (sauf cas particuliers en nombre limité) les rubriques relatives :

a) aux métiers

- de l'agriculture et des industries agricoles et alimentaires
- du bâtiment, travaux publics et topographie
- de la géologie, mines et carrières
- de la chimie
- des transports (sauf 513 et 514)
- des banques et des assurances, de la publicité et de l'information
- de l'enseignement et de la recherche, des activités littéraires, juridiques, artistiques et sportives,
- de la santé et des soins personnels
- des activités politiques, syndicales et religieuses.

b) aux emplois d'ouvriers qualifiés :

- conducteurs d'engins, de fours, de chaudières
- de l'extraction et du travail des pierres
- de la maçonnerie (sauf 320) et travaux connexes
- de la couverture, plomberie, installations sanitaires
- de la production et de la première transformation des métaux
- de la chimie, travail du verre, céramique, papier-carton, production des fibres artificielles et synthétiques
- de la photographie, impression, reliure,
- de l'alimentation
- des textiles, du travail des étoffes et des cuirs et peaux
- du travail du bois (sauf 463), des matières connexes et de la transformation des matières plastiques.

3) Les rubriques utilisées et l'analyse fonctionnelle

Les rubriques les plus fréquemment utilisées se rattachent :

a) A la direction et aux services administratifs

Désignation des emplois	Code E	Fréquence d'utilisation
Direction et cadres administratifs	100	47,15 %
	600	36,62 %
	601	15,91 %
	602	25,95 %
Comptables et caissiers	606	38,23 %
	607	22,71 %
Secrétaires, sténodactylographes, dactylographes	603	8,49 %
	609	14,09 %
	610	64,73 %
Autres personnels administratifs qualifiés	604	16,82 %
	605	7,68 %
	608	15,28 %
	611	29,05 %
Personnels administratifs sans qualification particulière	612	26,70 %

b) Aux services commerciaux

Désignation des emplois	Code E	Fréquence d'utilisation
Cadres supérieurs et techniques des services commerciaux	700	14,03 %
	701	19,25 %
Agents techniques des services commerciaux	702	8,29 %
	704	16,20 %
V.R.P.	703	12,74 %
Employés qualifiés des services commerciaux	707	5,32 %

c) Aux services généraux

Désignation des emplois	Code E	Fréquence d'utilisation
Maçons, peintres, menuisiers	320	8,69 %
	340	5,74 %
	463	9,42 %
Conducteurs poids lourds et véhicules automobiles	513	12,72 %
	514	22,26 %
Personnels <sup>de</sup> service et de gardiennage	951	11,00 %
	952	13,06 %

d) Aux services d'organisation de la production

Désignation des emplois	Code E	Fréquence d'utilisation
Ingénieurs et cadres techniques	203	15,61 %
	206	5,86 %
	223	8,29 %
	226	22,15 %
Techniciens de l'organisation du travail et des services de gestion de la production	250	17,81 %
	260	26,32 %
	261	9,85 %
Dessinateurs industriels	280	21,83 %
	284	34,66 %
	288	14,57 %
Autres techniciens	242	16,10 %

e) Aux services d'encadrement de la production

Désignation des emplois	Code E	Fréquence d'utilisation
Agents de maîtrise d'ouvriers qualifiés	271	59,27 %
	275	10,13 %
Chefs d'équipe d'ouvriers qualifiés	272	26,10 %
	276	11,57 %
Agents de maîtrise et chefs d'équipe d'O.S et manoeuvres	277	6,05 %

f) Aux services du magasinage et de la manutention

Personnel d'encadrement	500	21,25 %
Ouvriers et employés qualifiés	501	37,64 %
	503	5,94 %
Ouvriers et employés non qualifiés	502	31,26 %

g) Aux services d'entretien

Ouvriers qualifiés d'entretien	384	22,84 %
	391	18,77 %
	394	11,98 %

h) A la production

Désignation des emplois	Code E	Fréquence d'utilisation
Ouvriers qualifiés de la forge, de la soudure, de la serrurerie, de la chaudronnerie et de la charpente de fer	360	13,10 %
	361	38,44 %
	362	12,14 %
	370	19,15 %
	371	14,72 %
	372	7,98 %
Ouvriers qualifiés de la mécanique générale.	380	44,92 %
	382	30,69 %
Ouvriers qualifiés de la mécanique : ouvriers sur machines-outils	383	22,09 %
	387	74,77 %
	388	16,53 %
	389	5,68 %
Autres ouvriers qualifiés de la production	341	18,10 %
	470	9,21 %
Ouvriers spécialisés, manoeuvres	471	60,74 %
	472	62,42 %

Ainsi s'esquisse une possibilité de classification éventuelle des rubriques principales en quelques catégories fondamentales, qui, sans mettre parfaitement en évidence la notion de fonction, en présentent une approche partielle par l'intermédiaire de la notion de service. Il est certain que d'autres regroupements auraient aussi bien pu être réalisés, soit en jouant sur les proximités susceptibles d'exister entre le contenu de deux ou plusieurs rubriques, soit en utilisant un autre critère de classification comme celui de la qualification requise pour tenir chacun des emplois.



### III - LA STRUCTURE DECLAREE ET LA TAILLE DES ETABLISSEMENTS

#### 1) Les rubriques utilisées et la taille des établissements

Puisque les déclarations doivent fournir une répartition des salariés occupés, il est bien évident que le nombre de rubriques utilisées par un établissement quelconque pourra tout au plus être égal à l'effectif total qu'il emploie. Si l'on note en outre que les emplois correspondant à une activité professionnelle similaire doivent être regroupés sous une même rubrique, on conçoit qu'une telle contrainte devienne prépondérante lorsque la taille des établissements enquêtés est faible. Or, quatre établissements sur cinq étudiés occupent moins de 100 salariés ; un établissement sur cinq occupe moins de 20 salariés. En moyenne, deux rubriques absorbent à elles seules 30 % des effectifs ; six rubriques en absorbent près de 50 %.

De sorte que, si les petits établissements ont utilisé dans leur ensemble la même liste de rubriques que les grands (à quelques unités près), cela s'explique par une plus grande diversité dans le choix des rubriques retenues d'un établissement à l'autre. Diversité tenant soit à la division des tâches entre les établissements d'une même entreprise (unité productive pure dissociée des services administratifs et commerciaux, siège social, siège social et services commerciaux, etc), soit aux difficultés de classement des salariés par postes d'emploi (pluralité des tâches pour un même individu), soit enfin à une plus grande distorsion des modes organisationnels (suppression ou atrophie de certaines fonctions ou de certains emplois spécifiques dont l'existence est exceptionnelle dans les petits établissements).

Il en résulte que, d'une façon générale, la probabilité d'apparition d'une rubrique donnée dans une déclaration est d'autant plus faible que la taille de l'établissement est petite. En sens inverse,

certaines rubriques sont communes à la plupart des déclarations relatives aux grands établissements : 46 rubriques sont utilisées par plus de la moitié des établissements occupant 500 salariés et plus.

Si l'on se réfère à la liste des 66 emplois spécifiques les plus souvent retenus dans les déclarations, on obtient, à partir d'un découpage de la population globale en quatre classes (code T1), les distributions suivantes :

Nombre de rubriques utilisés par	$N < 20$	$20 \leq N \leq 99$	$100 \leq N \leq 499$	$N \geq 500$
Moins de 1 % des établissements	4	-	-	-
1 à 4 % des établissements	27	8	-	-
5 à 9 % des établissements	19	11	-	-
10 à 14 % des établissements	4	19	4	-
15 à 19 % des établissements	4	7	7	3
20 à 29 % des établissements	1	9	13	4
30 à 49 % des établissements	6	7	17	13
50 à 74 % des établissements	1	5	18	20
75 à 90 % des établissements	-	-	6	15
Plus de 90 % des établissem.	-	-	1	11
Total	66	66	66	66

On assiste dès lors à une translation générale de la distribution des fréquences d'utilisation, en fonction de la taille des établissements. C'est ainsi que, dans la sous-population des petits établissements (taille inférieure à 20), 50 des 66 rubriques retenues sont utilisées par moins d'un établissement sur dix ; une seule par plus d'un établissement sur deux ; aucune par plus de trois établissements sur quatre. En revanche, dans la sous-population des grands établissements (taille supérieure à 500), 59 rubriques sont utilisées

par plus de trois établissements sur dix ; 11 par plus de neuf établissements sur dix ; aucune par moins de 15 % des établissements.

Le choix d'un seuil de fréquence à 5 %, dicté par l'observation de la distribution globale, a ainsi conduit à dégager une liste d'emplois spécifiques qui, pour l'essentiel, apparaissent :

- dans plus de 30 % des déclarations relatives aux établissements occupant plus de 500 salariés (2,5 % de la population).
- dans plus de 15 % des déclarations relatives aux établissements occupant 100 à 500 salariés (15,5 % de la population) ;
- dans plus de 5 % et moins de 75 % des déclarations relatives aux établissements occupant 20 à 100 salariés (56 % de la population)
- dans plus de 1 % et moins de 50 % des déclarations relatives aux établissements occupant moins de 20 salariés (26 % de la population).

L'instrument d'analyse mis en oeuvre, privilégiant le critère de la fréquence d'apparition de chacune des rubriques, exprime donc le résultat d'un effet de composition de la population par taille. La prédominance des petits établissements explique en grande partie que, même les emplois les plus fondamentaux, qui a priori devraient exister dans la majorité des établissements, ne se retrouvent généralement pas dans un grand nombre de déclarations.

Deux questions doivent alors être posées :

- les divergences observées dans le choix des rubriques, principalement pour les petits établissements, reflètent-elles nécessairement une grande distorsion des modes organisationnels ?
- l'accroissement des fréquences d'utilisation des rubriques en fonction de la taille des établissements, ne doit-il pas mettre en évidence des catégories d'emploi plus spécifiques aux grandes unités productives ?

FREQUENCE D'UTILISATION DES PRINCIPALES RUBRIQUES UTILISEES (I)

Services	Désignation des emplois	Liste des emplois
SERVICES  ADMINIS-  TRATIFS	Direction et cadres administratifs	Chef d'entreprise Directeur d'établissement Cadres administratifs supérieurs Autres cadres administratifs
	Comptables et caissiers	Personnel qualifié des services comptables Aide-comptables, teneurs de livres, encaisseurs
	Secrétaires, sténodactylographes, dactylographes	Secrétaires de direction Secrétaires spécialisées Personnel de secrétariat
	Autres personnels administratifs qualifiés	Personnel d'encadrement Employés qualifiés des services administratifs et contentieux Standardistes Autres employés de bureau qual.
	Employés non qualifiés	Employés de bureau non qual.
SERVICES  COMMERCIAUX	Cadres supérieurs et techniques.	Cadres supérieurs des serv.com. Cadres techniques des serv.com.
	Agents techniques	Personnel tech. des serv.achat Personnel tech. serv.vente (non V.R.P)
	V.R.P.	V.R.P.
	Employés qualifiés	Personnel qualifié de la vente
SERVICE DU TRAITEMENT  DE L'INFORMATION	Analystes, programmeurs	Analystes Programmeurs
	Employés qualifiés	Opérateurs Calculateurs et calc.mécanographes. Perforeurs, perforeurs-vérificateurs

(suite I)

Code E	Toutes % Ets	tailles % Eff.	N < 20 % Ets	20 < N % Eff.	20 < N % Ets	< 99 % Eff.	100 < N % Ets	< 499 % eff.	N > 500 % Ets	500 % Eff.
100	47,15	0,62	42,17	3,08	49,42	1,26	48,62	0,27	38,19	0,04
600	36,62	0,53	26,06	2,01	36,27	0,99	52,01	0,33	58,27	0,08
601	15,91	0,36	3,44	0,26	10,40	0,32	46,11	0,41	80,71	0,36
602	25,95	0,62	7,87	0,59	23,81	0,74	55,78	0,63	75,59	0,47
606	38,23	1,13	15,58	1,18	35,66	1,14	76,38	1,21	92,91	1,01
607	22,71	0,66	7,22	0,56	19,10	0,65	52,76	0,69	77,17	0,67
603	8,49	0,12	4,58	0,32	7,77	0,20	15,83	0,11	19,29	0,04
609	14,09	0,37	7,90	0,64	11,78	0,43	27,76	0,39	44,88	0,25
610	64,73	3,22	41,30	4,01	66,73	3,70	91,21	3,40	97,64	3,27
604	16,82	0,74	3,40	0,26	11,59	0,42	46,17	0,78	90,55	1,09
605	7,68	0,29	2,38	0,18	4,69	0,16	19,66	0,33	55,12	0,40
608	15,28	0,23	0,98	0,07	7,62	0,19	54,90	0,32	88,98	0,17
611	29,05	1,16	16,19	1,37	26,19	1,13	54,15	1,37	70,08	0,87
612	26,70	1,50	8,13	0,68	22,57	1,00	62,88	1,82	86,61	1,71
700	14,03	0,30	5,41	0,41	11,40	0,37	32,29	0,30	48,82	0,21
701	19,25	0,85	6,58	1,02	16,31	0,97	44,85	0,99	57,09	0,49
702	8,29	0,23	1,06	0,10	5,12	0,18	26,01	0,27	44,09	0,26
704	16,20	0,69	7,19	0,94	14,94	0,83	31,41	0,74	43,31	0,43
703	12,74	0,65	8,28	1,33	14,06	1,12	14,95	0,55	15,35	0,19
707	5,32	0,22	2,12	0,25	4,50	0,24	11,62	0,22	17,72	0,20
620	1,61	0,03	0,23	0,02	0,38	0,01	3,64	0,03	31,10	0,06
621	1,76	0,04	0,08	0,01	0,21	0,01	4,46	0,04	37,40	0,09
622	2,44	0,07	-	-	0,40	0,01	8,29	0,08	37,40	0,13
623	4,50	0,12	0,49	0,04	2,49	0,09	14,64	0,15	27,95	0,12
624	3,05	0,16	-	-	0,30	0,02	10,74	0,19	48,82	0,28

FREQUENCE D'UTILISATION DES PRINCIPALES RUBRIQUES UTILISEES (II)

Services	Désignation des emplois	Liste des emplois
SERVICE A CARACTERE SOCIAL	Médecin, infirmière, assistante sociale	Médecins Auxiliaires médicaux Personnel diplômé des services sociaux
SERVICES GENERAUX	Maçons, peintres, menuisiers	Maçons Peintres en bâtiment Ouvriers qualifiés de la menuiserie
	Conducteurs de poids lourds et véhicules automobiles	Agents de conduite de transport routier Conducteurs de véhicules automobiles
	Personnel de service et de gardiennage	Personnel de gardiennage Autre personnel de service
ORGANISATION DE LA PRODUCTION	Ingénieurs et cadres techniques	Ingénieurs spécialistes des sciences physiques Ingénieurs mécaniciens Cadres techniques spécialistes des sciences physiques Cadres techniques spécialistes de la mécanique
	Techniciens de l'organisation du travail et des services de gestion de la production	Techniciens de l'organisation du travail Techniciens des services de gestion de la production Techniciens du contrôle de la gestion de la production
	Dessinateurs industriels	Chef de groupe, projeteurs et dessinateurs principaux Dessinateurs d'études et de petites études Dessinateurs d'exécution
	Autres techniciens	Techniciens des services des essais, contrôles, laboratoires
ENCADREMENT DE LA PRODUCTION	Agents de maîtrise d'ouvriers qualifiés	A.M d'O.Q. du travail des métaux et de l'électricité A.M d'O.Q autres spécialités
	Chef d'équipe d'ouvriers qualifiés	C.E d'OQ du tr.métaux et élect. C.E d'O.Q autres spécialités
	A.M et C.E d'O.S et O.M	A.M et C.E d'OS et O.M

(Suite II)

Code E	Toutes tailles		N < 20		20 ≤ N ≤ 99		100 ≤ N ≤ 499		N > 500	
	% Ets	% Eff.	% Ets	% Eff.	% Ets	% Eff.	% Ets	% Eff.	% Ets	% E
900	1,36	0,02	-	-	0,12	-	3,02	0,01	33,07	0,04
904	3,73	0,06	-	-	0,03	-	12,63	0,06	70,08	0,11
941	1,75	0,02	0,04	-	0,17	-	4,33	0,02	38,98	0,04
320	8,69	0,25	1,63	0,18	4,79	0,21	25,00	0,27	68,11	0,29
340	5,74	0,16	0,79	0,06	3,69	0,13	15,26	0,19	43,70	0,17
463	9,42	0,49	1,21	0,17	5,86	0,26	26,51	0,36	68,11	0,95
513	12,72	0,29	2,69	0,26	10,84	0,36	31,28	0,31	43,31	0,19
514	22,26	0,41	6,35	0,46	22,03	0,61	41,58	0,36	72,05	0,26
951	11,00	0,31	0,61	0,04	5,98	0,16	36,56	0,33	72,44	0,46
952	13,06	0,29	7,30	0,51	13,61	0,40	19,22	0,26	22,05	0,18
203	15,61	0,64	3,78	0,38	11,15	0,43	42,02	0,78	74,02	0,69
206	5,86	0,18	1,74	0,14	4,78	0,17	12,69	0,18	30,31	0,22
223	8,29	0,39	1,66	0,14	5,09	0,24	23,99	0,45	51,18	0,50
226	22,15	0,59	11,54	0,91	21,70	0,76	37,06	0,51	49,61	0,46
250	17,81	1,02	2,46	0,18	11,48	0,40	54,40	1,10	91,34	1,66
260	26,32	1,23	4,50	0,33	21,38	0,74	69,41	1,31	94,88	1,76
261	9,85	0,34	3,21	0,25	8,26	0,27	21,55	0,26	41,34	0,55
280	21,83	1,00	5,14	0,50	17,31	0,79	56,16	1,19	82,68	1,04
284	34,66	1,55	11,12	1,04	33,13	1,48	71,04	1,85	86,22	1,31
288	14,57	0,64	1,89	0,15	10,00	0,37	41,77	0,82	79,13	0,75
242	16,10	0,89	3,44	0,40	11,54	0,47	44,41	1,01	73,62	1,23
271	59,27	2,20	40,73	3,33	60,47	2,59	80,72	2,06	90,55	1,81
275	10,13	0,27	6,28	0,51	10,18	0,40	13,25	0,20	29,53	0,20
272	26,10	1,52	9,53	0,86	23,58	1,35	54,77	1,75	75,98	1,49
276	11,57	0,53	4,16	0,35	11,31	0,56	21,61	0,56	31,50	0,49
277	6,05	0,37	1,44	0,14	3,50	0,15	15,58	0,30	51,97	0,71

FREQUENCE D'UTILISATION DES PRINCIPALES RUBRIQUES UTILISEES (III)

Services	Désignation des emplois	Liste des emplois
MAGASINAGE ET MANUTENTION	Personnel d'encadrement	Personnel d'encadrement
	Personnel qualifié	Ouvriers et empl. qualifiés Conducteurs d'appareils de levage.
	Personnel non qualifié	Ouvriers et employés non qualifiés
ENTRETIEN	Ouvriers qualifiés d'entretien	Ouvriers mécaniciens d'entret. Electriciens d'entretien Electromécaniciens d'entretien
EXECUTION DE LA PRODUCTION	Ouvriers qualifiés de la forge, de la soudure de la serrurerie et de la charpente de fer	Ouvriers qualifiés du travail à chaud des pièces mét. Soudeurs professionnels Serruriers et assimilés Chaudronniers et traceurs en chaudronnerie Tôliers Charpentiers en fer et assim.
	Ouvriers qualifiés de la mécanique générale	Ajusteurs et assimilés Ajusteurs-monteurs-mécaniciens
	Ouvriers qualifiés sur machines-outils	Ajusteurs - outilleurs Ouvriers profes. sur machines Ouvriers profes. sur machines, hautement qualifiés Régleurs de machines-outils automatiques
	Autres ouvriers qualifiés	Ouvriers qual. dans les travaux de peinture indust. Autres ouvriers qualifiés
	Ouvriers spécialisés, manoeuvres.	Ouvriers spécialisés Manoeuvres



(Suite III)

Code E	Toutes tailles		N < 20		20 ≤ N ≤ 99		100 ≤ N ≤ 499		N ≥ 500	
	% Ets	% Eff	% Ets	% Eff.	% Ets	% Eff.	% Ets	% Eff.	% Ets	% Eff
500	21,25	0,47	3,37	0,24	16,49	0,47	58,29	0,58	82,68	0,37
501	37,64	1,47	10,40	0,81	37,87	1,32	74,06	1,51	87,80	1,65
503	5,94	0,37	0,19	0,02	2,00	0,08	20,73	0,39	62,20	0,67
502	31,26	2,24	9,27	0,86	28,58	1,39	67,96	2,40	90,55	3,07
384	22,84	1,42	9,27	1,55	18,68	1,28	51,70	1,35	77,17	1,65
391	18,77	0,74	2,42	0,25	12,46	0,47	58,10	0,87	85,04	0,93
394	11,98	0,63	4,20	0,61	9,64	0,62	27,64	0,70	47,64	0,55
360	13,10	0,45	3,10	0,33	9,46	0,47	33,35	0,47	72,44	0,42
361	38,44	4,07	19,06	3,03	37,75	4,03	64,70	3,92	91,34	4,48
362	12,14	0,96	6,16	1,25	10,87	1,16	20,29	0,91	51,97	0,80
370	19,15	2,48	8,21	1,60	16,78	2,27	37,44	2,58	72,05	2,70
371	14,72	1,12	5,86	0,89	12,58	1,24	29,40	1,13	63,39	1,03
372	7,98	1,56	2,72	0,88	7,27	1,08	16,27	1,72	26,77	1,94
380	44,92	4,00	36,50	6,42	45,42	5,44	53,64	3,57	66,54	2,75
382	30,69	4,18	15,92	2,97	28,13	3,70	57,16	4,38	76,38	4,58
383	22,09	1,61	12,97	2,95	18,26	1,92	42,40	1,25	75,98	1,56
387	74,77	12,56	65,77	22,70	74,80	17,13	86,37	11,62	94,88	7,64
388	16,53	1,50	9,83	2,59	14,17	1,92	28,20	1,16	66,54	1,40
389	5,68	0,40	1,10	0,12	3,38	0,18	16,58	0,43	37,01	0,62
341	18,10	0,82	4,35	0,33	15,41	0,61	42,71	0,85	67,72	1,05
470	9,21	1,11	2,04	0,33	7,18	1,00	23,05	0,97	42,91	1,52
471	60,74	16,56	39,49	9,34	62,34	13,37	84,55	16,65	96,46	20,80
472	62,42	4,29	41,30	5,59	64,64	5,72	84,92	4,12	90,94	2,89

2) La répartition des emplois et la taille des établissements

Lorsque la taille des établissements enquêtés augmente, l'analyse des fréquences d'utilisation de chacune des rubriques semble devoir s'orienter dans quatre directions conjointes. On observe en effet :

- un phénomène de spécialisation accrue des tâches productives, qui constitue la tendance générale;
- un accroissement de la division des tâches par structuration fonctionnelle et structuration hiérarchique plus prononcées;
- un développement des services autonomes à compétence technique;
- l'apparition de nouveaux services spécialisés.

a) La spécialisation accrue des tâches productives

Le résultat n'est plus à démontrer : il se dégage spontanément des chiffres précédents. Toutefois à titre d'illustration, le tableau ci-dessous précise l'ampleur du phénomène. Nous avons sélectionné les sept rubriques regroupant le plus grand nombre de salariées dans les petits établissements (à l'exception cependant des ouvriers spécialisés, le poids relatif tend à augmenter avec la taille des établissements, conformément au processus de spécialisation accrue).

- Pourcentage des effectifs employés -

Code E	$N < 20$	$20 \leq N \leq 99$	$100 \leq N \leq 499$	$N \geq 500$
100	3,08 %	1,26 %	0,27 %	0,04 %
271	3,33 %	2,59 %	2,06 %	1,81 %
380	6,42 %	5,44 %	3,57 %	2,75 %
387	22,70 %	17,33 %	11,62 %	7,64 %
472	5,59 %	5,72 %	4,12 %	2,89 %
600	2,01 %	0,99 %	0,33 %	0,08 %
610	4,01 %	3,70 %	3,40 %	2,37 %
Total	47,14 %	37,03 %	25,37 %	17,58 %

Les sept rubriques principales utilisées par les petits établissements regroupent à elles seules près de la moitié des effectifs de la sous-population. Lorsque la taille des unités productives augmente, le poids relatif des catégories d'emplois correspondant à ces rubriques tend à décroître fortement : il passe à moins de 20% dans les plus grands établissements. Corrélativement, les effectifs des 59 autres rubriques retenues le plus souvent dans les déclarations, augmentent sensiblement en valeur relative : ils représentent moins de 50% du total des salariés occupés dans les plus petits établissements et plus de 75% du total des salariés occupés dans les plus grands établissements. L'observation du poids des 135 rubriques restantes (95 rubriques ne sont jamais utilisées) confirme encore, bien que faiblement, ce phénomène tendanciel.

- Pourcentage des effectifs employés -

	N < 20	20 ≤ N ≤ 99	100 ≤ N ≤ 499	N ≥ 500
Liste 1 (7 rubriques)	47,14 %	37,03 %	25,37 %	17,58 %
Liste 2 (59 rubr.)	49,07 %	59,18 %	69,70 %	76,17 %
Sous-total (66 rubr.)	96,21 %	96,21 %	95,07 %	93,75 %
Liste 3 (135 rubr.)	3,79 %	3,79 %	4,93 %	6,25 %
Total (201 rubriques)	100,00 %	100,00 %	100,00 %	100,00 %

Ainsi, bien que le nombre global de rubriques utilisées varie peu d'une classe de taille à l'autre, le poids représenté par l'effectif de chacune de ces rubriques diffère sensiblement selon la dimension de l'unité productive. Au-dessus d'une certaine taille, les spécialisations se multiplient. En revanche, dans les petits établissements, la spécialisation est faible, l'ensemble des tâches étant effectuées par un petit nombre de salariés. L'affectation à une rubrique procède alors, semble-t-il, d'une appréciation de la fonction principale de l'individu, indépendamment des autres fonctions qu'il est susceptible d'assumer. Le principe même de la classification des emplois, privilégiant un des aspects des activités individuelles au détriment des autres déforme le contenu réel de la notion de structure chaque fois que le cadre organisationnel repose sur une combinaison de tâches intégrées, accomplies par un petit nombre de salariés aux compétences polyvalentes.

b) Développement de services autonomes à compétence technique

Il s'agit principalement des services suivants :

- organisation du travail et gestion de la production,
- dessinateurs industriels,
- entretien
- magasinage et manutention.

En règle générale, s'ils ne sont pas systématiquement absents des déclarations relatives aux petits établissements, ces divers services n'y apparaissent qu'exceptionnellement, consacrant à ce niveau une grande confusion des fonctions dont la structuration n'est évidente qu'au delà d'une certaine taille. Les spécialités que nous avons isolées ici (1) existent de façon autonome :

- dans moins d'un établissement sur cinq occupant 10 à 20 salariés ;
- dans moins d'un établissement sur deux occupant 20 à 100 salariés ;
- dans plus de huit établissements sur dix occupant 100 à 500 salariés ;
- quasiment dans tous les établissements occupant plus de 500 salariés.

Encore faut-il souligner que dans les petits établissements, la hiérarchie des niveaux de qualification internes aux services, est le plus souvent écrasée lorsqu'elle ressort de la nomenclature (dessinateurs industriels, personnel du magasinage et manutention), la fonction pouvant à la limite être représentée par une seule catégorie de salariés. L'observation simultanée des techniciens de l'organisation du travail et des techniciens de la gestion de la production, ne se manifeste que dans les établissements occupant plus de 100 salariés. Il en va de même de la distinction entre ouvriers mécaniciens et ouvriers électro mécaniciens d'entretien.

---

(1) Les chiffres indiqués correspondent à la fréquence d'apparition de l'une quelconque des rubriques rattachées au service. Toutefois, dans le cas des dessinateurs industriels, nous avons pris en compte l'ensemble du personnel (sous-groupe 28), quelle que soit sa spécialité (mécanique, électricité et électronique, bâtiment, architecture, travaux publics et autres spécialités). En ce qui concerne le service du magasinage et manutention, seul le personnel qualifié a été décompté (sous-groupe 50, à l'exclusion de la rubrique 502) excluant ainsi le personnel non qualifié souvent confondu avec les manoeuvres et ouvriers spécialisés.

Services et rubriques	Fréquences d'utilisation			
	N < 20	20 ≤ N < 99	100 ≤ N < 499	N ≥ 500
1. ORGANISATION DU TRAVAIL ET GESTION DE LA PRODUCTION	9,90 %	33,72 %	81,15%	97,63%
- Préparateurs et techniciens de l'organisation du travail de production et des postes de travail (250)	2,46 %	11,48%	54,40%	91,34%
- Personnel technique des services de gestion de la production (260)	4,50%	21,38%	69,41%	94,88%
- Personnel assurant le contrôle de la gestion de la production (261)	3,21%	8,26%	21,55%	41,34%
2. DESSINATEURS INDUSTRIELS	17,24%	47,14%	87,68%	98,42%
- Chef de groupe, projeteurs et dessinateurs principaux (mécanique) : 280	5,14 %	17,31%	56,16%	82,68%
- Dessinateurs d'études et de petites études (mécanique) : 284	11,12%	33,13%	71,04%	86,22%
- Dessinateurs d'exécution, dessinateurs catalogues, dessinateurs détaillants : 288	1,89%	10,00%	41,77%	79,13%
3. ENTRETIEN	14,93%	34,66%	84,61%	97,63%
- Ouvriers mécaniciens d'entretien(384)	9,27%	18,68%	51,70%	77,17%
- Monteurs électriciens (391)	2,42%	12,46%	58,10%	85,04%
- Mécaniciens électriciens, mécaniciens électroniciens, électromécaniciens d'entretien (394)	4,20%	9,64%	27,64%	47,64%
4. MAGASINAGE ET MANUTENTION	13,61%	49,04%	89,51%	96,85%
- Personnel d'encadrement	3,37%	16,49%	58,29%	82,68%
- Ouvriers et employés qualifiés	10,40%	37,87%	74,06%	87,80%
- Ouvriers et employés non qualifiés	9,27%	28,58%	67,87 %	90,55%

c) Apparition de nouveaux services spécialisés

Il s'agit principalement :

- du service de traitement électromécanique et électronique de l'information (1);
- des services à caractère social (conçus au sens large).

Ces services sont généralement inexistantes dans les établissements de petite taille. Dans les entreprises de dimension moyenne, leur apparition est encore exceptionnelle. Seules les grandes unités productives semblent en mesure de rentabiliser les techniques de gestion moderne (ordinateur) et d'entretenir des catégories de personnel à compétence socio-médicale.

Fréquences d'utilisation

Désignation du service	Désignation des emplois	Code E	$N < 20$	$20 \leq N \leq 99$	$100 \leq N \leq 499$	$N \geq 500$
Traitement électromécanique et électronique de l'information	Analystes-programmeurs	620	0,23%	0,38%	3,64%	31,10
		621	0,08 %	0,21%	4,46%	37,40%
		620-621	0,30%	0,59%	6,34%	44,88%
Opérateurs sur ordinateur, mécatronographes, calculat., perforeurs		622	-	0,40%	8,29%	37,40%
		623	0,49%	2,49%	14,64 %	27,95%
		624	-	0,30%	10,74%	48,82%
		622 à 624	0,49%	2,91%	23,61%	62,20%
Service à caractère social	Médecin, infirmière, assistante sociale.	900	-	0,12%	3,02%	33,07
		904	-	0,03%	12,63%	70,08%
		941	0,04	0,17%	4,33%	27,95

(1) On remarquera qu'une partie des cadres responsables du traitement de l'information est susceptible de figurer sous les rubriques 207, 217, ou 227 (ingénieurs et cadres techniques, spécialistes, entre autres, de l'informatique). Mais, en sens inverse, ces rubriques ne peuvent être rattachées au service informatique en raison de la multiplicité des attributions qu'elles tendent à recouvrir (recherche, méthode, bureau calcul, organisation de la production, gestion administrative, etc...)

On retrouve à ce niveau, le défaut de l'analyse globale, dont les résultats, tributaires d'un effet de composition de la population par taille, risquaient de marquer l'importance de certaines catégories d'emplois spécifiques à une fraction des établissements. Il est clair que ces emplois ne peuvent être négligés, même si leur apparition est en règle générale exceptionnelle, leur absence ou leur présence dans les déclarations constituant un indicateur du mode organisationnel choisi. Nous obtenons donc en définitive une liste de 74 rubriques (66 initiales + 8 nouvelles) susceptibles de constituer un cadre de référence à l'analyse structurelle des emplois dans les établissements de la mécanique.

d) Structuration fonctionnelle et hiérarchique plus prononcée

Petits établiss. (N < 20) : rubriques utilisées par :	Grands établissements (N > 500)										Total
	Rubriques utilisées par :										
	1%	1 à 4%	5 à 9%	10 à 14%	15 à 19%	20 à 29%	30 à 49%	50 à 74%	75 à 89%	90 à 100%	
Moins de 1 %						1	7	3	1		12
1 à 4 %					2	1	8	9	4	3	27
5 à 9 %					1	2	3	6	6	1	19
10 à 14 %							1		3		4
15 à 19 %								1	1	2	4
20 à 29 %								1			1
30 à 49 %							1	1		4	6
50 à 74 %										1	1
75 à 89 %											
90 à 100 %											
Total					3	4	20	21	15	11	74

L'apparente simplicité des résultats qui se dégage du tableau précédent ne doit pas faire illusion. S'il apparaît nettement une relation entre la dimension de l'unité productive et la fréquence d'utilisation de chacune des rubriques, n'est-ce pas en grande partie parce que le contenu même des emplois change d'un établissement à l'autre ? La définition d'une rubrique se réfère aux deux aspects complémentaires des emplois qu'elle désigne, c'est-à-dire :

- aux tâches élémentaires susceptibles d'être rattachées à la nature des emplois désignés
- et aux qualités et capacités requises pour tenir ces emplois.

Or, le nombre et surtout l'importance et la complexité des tâches élémentaires n'apparaissent pas comme une donnée immuable indépendante de l'établissement. La correspondance établie, qui sert de support à la définition de la rubrique, ne constitue dès lors qu'une approximation, consacrant seulement un niveau d'équilibre possible entre un processus de division et un processus d'intégration des tâches.

Les conséquences en sont multiples. C'est ainsi qu'une rubrique peut regrouper des emplois assez nettement différenciés si la spécialisation est grande. Mais l'on peut toujours considérer dans ce cas que le regroupement s'appuie sur l'existence d'un caractère commun que l'on s'est efforcé de mettre en évidence. En revanche, si la spécialisation est faible, un emploi, défini par son titulaire, est susceptible de recouvrir une gamme relativement étendue de tâches différentes et par suite de déborder le cadre de définition de la rubrique; à la limite, seule l'utilisation conjointe de plusieurs rubriques pourrait convenir. Mais il serait nécessaire dans ce cas de disposer d'une nomenclature établie selon un principe d'emboitage, qui, en reposant sur une hiérarchie des proximités observées, permettrait de relier les rubriques entre-elles.

La nomenclature des emplois à laquelle se réfèrent les déclarations, n'a pas été établie selon ce principe. Le choix des diverses rubriques retenues, lié au problème du rapprochement entre découpage



réel des tâches et modèle préétabli, est subordonné lui-même au choix des critères permettant de réaliser l'adéquation. Or, ces critères sont mal connus, et, si, en règle générale, ils consacrent le principe de l'activité principale exercée, leur imprécision reste grande. Comment concilier en effet le temps d'exécution de chacune des tâches élémentaires, leur complexité relative, les qualités et capacités requises pour accomplir toutes les étapes du processus de travail, sans introduire une part déterminante d'appréciation subjective ?

On sait seulement que le problème des affectations multiples est nécessairement résolu : la classification d'un emploi sous une rubrique donnée exclut sa classification sous une autre rubrique. Un tel principe d'exclusion admet deux conséquences. Si le processus de production repose sur une grande division des tâches, le nombre de rubriques utilisées simultanément sera grand, et d'une façon générale, l'utilisation d'une rubrique impliquera l'utilisation de l'ensemble des autres. En sens inverse, si la spécialisation est faible, les rubriques utilisées simultanément seront peu nombreuses, et d'une façon générale, l'utilisation de certaines rubriques impliquera l'exclusion de certaines autres rubriques et inversement. Dans le premier cas, on débouche sur la notion de complémentarité de l'ensemble des rubriques entre elles ; dans le deuxième cas sur la notion de substituabilité de deux ou de plusieurs rubriques entre-elles.

Ainsi semblent devoir s'expliquer les résultats dégagés. Au-dessous d'une certaine taille, la spécialisation est faible : les rubriques utilisées sont peu nombreuses et elles tendent à s'exclure mutuellement au sein de "familles d'emplois substituables" dont l'existence consacre l'intégration possible des tâches. Au-dessus d'une certaine taille, les entreprises opèrent des spécialisations : la liste des rubriques utilisées tend à s'allonger et progressivement apparaissent des "familles d'emplois complémentaires" dont la combinaison consacre un phénomène de structuration fonctionnelle et hiérarchique de plus en plus prononcé.

Ainsi semble devoir également s'expliquer l'inadaptation partielle du système de représentation traditionnel des emplois sous la forme d'une simple liste d'appellations exclusives les unes des autres, négligeant l'influence combinée du double phénomène de distance et de proximité entre emplois, support nécessaire à une analyse du phénomène plus général de comparaison des structures entre elles.

*C H A P I T R E   I I*

---

*COMPARAISON STRUCTURELLE ET ANALYSE QUALITATIVE*

*RECHERCHE EMPIRIQUE*

## I - REPRESENTATION STRUCTURELLE ET DISTANCE ENTRE EMPLOIS

### 1. Les critères qualitatifs

La nomenclature des emplois repose sur une combinaison fondamentale de trois critères principaux : la qualification requise, le service d'appartenance et la spécialité exercée. Il est possible notamment de distinguer cinq niveaux de qualification successifs : les ingénieurs et cadres ; les techniciens et agents techniques ; le personnel d'encadrement ; le personnel qualifié ; le personnel sans qualification particulière. Nous avons vu précédemment, en analysant le contenu des rubriques le plus souvent utilisées dans les déclarations, qu'un regroupement des emplois en une dizaine de "services" semblait pouvoir être esquissé. Il s'agissait du service administratif, des services commerciaux, du service de traitement de l'information, des services à caractère social, des services généraux, du service d'organisation de la production, du service d'encadrement de la production, du service du magasinage et manutention, du service entretien et du service d'exécution de la production. En ce qui concerne la spécialité exercée, nous citerons notamment l'éclatement possible des employés de bureau qualifiés en comptables et caissiers, en secrétaires, sténodactylographes et dactylographes, et en autres personnels administratifs qualifiés ; des agents techniques des services commerciaux en VRP et autres agents techniques ; des techniciens de l'organisation de la production en techniciens de l'organisation du travail et des services de gestion de la production, en dessinateurs industriels et en autres techniciens ; des ouvriers qualifiés de la production en ouvriers qualifiés de la forge, de la soudure, de la serrurerie, de la chaudronnerie et de la charpente de fer, en ouvriers qualifiés de la mécanique générale, en ouvriers qualifiés sur machines-outils et en autres ouvriers qualifiés.

Ainsi semblerait s'ébaucher la construction d'un système d'emboîtement des postes de la nomenclature, pouvant conduire à regrouper par degrés successifs les diverses rubriques entre-elles. En pratique, il

n'existe aucune raison a priori permettant de hiérarchiser notamment les critères de qualification et de service. Si l'on adopte un passage donné des rubriques spécifiques en grandes spécialités (regroupement des 74 rubriques principales en 28 spécialités par exemple), la poursuite du système d'emboîtement peut suivre deux chemins différents selon le critère privilégié, comme le montrent les tableaux des pages suivantes. Nous considérerons dès lors que le croisement qualification - service fournit un découpage structurel minimal. On obtient ainsi une liste de 19 catégories d'emplois regroupés qui constituent une répartition de base fondamentale susceptible de coiffer la ventilation du personnel salarié entre les diverses rubriques de la nomenclature. Cette liste fondamentale se compose des postes suivants :

- cadres administratifs supérieurs
- employés qualifiés des services administratifs
- employés de bureau sans qualification particulière
- cadres supérieurs et techniques des services commerciaux
- agents techniques des services commerciaux
- employés qualifiés des services commerciaux
- cadres supérieurs et techniques du traitement de l'information,
- employés qualifiés du traitement de l'information
- personnel des services à caractère social
- personnel qualifié des services généraux
- personnel des services généraux sans qualification particulière
- ingénieurs et cadres techniques
- techniciens de la production
- personnel d'encadrement de la production
- personnel qualifié du magasinage et de la manutention
- personnel non qualifié de magasinage et de la manutention
- ouvriers qualifiés d'entretien
- ouvriers qualifiés de la production
- ouvriers sans qualification particulière.

PROPOSITION D'UNE GRILLE HIERARCHISEE PRIVILEGIANT LE CRITERE DE "SERVICE" PAR RAPPORT A CELUI DE "QUALIFICATION"

	Service	Qualification	Spécialité	Rubr.		
S E R V I C E	SERVICE	Cadres administratifs supérieurs	Direction et cadres administratifs	100		
				600		
	ADMINISTRATIF	Employés de bureau qualifiés			601	
					602	
					Comptables et caissiers	606
					607	
					Secrétaire, sténo-dactylographes, dactylographes.	603
	609					
	610					
	604					
	605					
	608					
	611					
	Employés non qualifiés	Empl. bureau non qualifiés	612			
I C E	SERVICES COMMERCIAUX	Cadres supérieurs et techniques	Cadres supérieurs et tech. serv.com.	700		
				701		
		Agents techniques			702	
					704	
					V.R.P.	703
	Employés qualifiés	Empl. com. qualifiés	707			
SERVICE DU TRAITEMENT DE L'INFORMATION	Cadres supérieurs et techniques	Analystes programmeurs		620		
				621		
SERVICE A CARACTERE SOCIAL	Assimilés cadres	Médecin, infirmière assistante sociale		622		
				623		
				624		
SERVICES GENERAUX	Personnel qualifié	Maçons; peintres, menuisiers.		320		
				340		
		463				
	Personnel sans qualif. particulière	Personnels de service et de gardiennage		513		
				514		
				951		
				952		

P R O D U C T I O N	ORGANISATION DE LA PRODUCTION	Ingénieurs et cadres techniques	Ingénieurs et cadres techniques	203	
				206	
		Techniciens	Techniciens de l'organisation du travail et des serv.de gestion de la production.	223	
				226	
			250		
	ENCADREMENT DE LA PRODUCTION	Personnel d'encadrement	Agents de maîtrise et chefs d'équipe	Dessinateurs industriels	260
					261
				Autres techniciens	280
	MAGASINAGE ET MANUTENTION	Personnel qualifié	Ouvriers et emp. Qual. du magasinage et de la manutention		284
					288
	ENTRETIEN	Personnel non qualifié	Manutentionnaires non qualifiés		242
					271
	U C T I O N	EXECUTION DE LA PRODUCTION	Ouvriers qualifiés	Ouvriers mécan., électriciens, électromécan. d'entretien	275
					272
		Ouvriers qualifiés	Ouvriers qualifiés	Ouvriers qualifiés de la forge, soudure, chaudronnerie et de la charpente de fer.	
					277
					500
					501
					503
					502
					384
					391
	394				
Ouvriers qualifiés	Ouvriers qualifiés	Ouvriers qualifiés sur machines-outils		360	
				361	
				362	
				370	
Ouvriers sans qualification particulière	Ouvriers sans qualification particulière	Ouvriers spécialisés manoeuvres.		371	
				372	
		Autres ouvriers qualifiés	380		
			382		
			383		
			387		
			388		
			389		
			341		
			470		
			471		
			472		

PROPOSITION D'UNE GRILLE HIERARCHISEE PRIVILEGIANT LE  
CRITERE DE "QUALIFICATION" PAR RAPPORT A CELUI DE "SERVICE"

	Qualification	Service	Spécialité	Rubr	
G E S T I O N	INGENIEURS	Services adminis- tratifs	Direction et ca- dres administratifs	100	
				600 601 602	
	ET	Services commerciaux	Cadres supérieurs et techniques des serv.commerciaux	700	
				701	
	CADRES	Serv.du traitement de l'information	Analystes programmeurs	620	
				621	
				900 904 941	
			Service à caractère social	Médecin, infirmière assist. sociale	203
					206 223
		TECHNICIENS	Services commerciaux	Agents techniques des serv.Commerciaux	702
704					
ET				V.R.P.	
				703	
AGENTS		Organisation de la production	Techniciens de l'or- ganisation du travail et serv.de gestion de production	250	
TECHNIQUES					260 261
					280 284 288
			Autres techniciens	242	



N O N I T I U C E X E	PERSONNEL D'ENCADRE- MENT	Encadrement de la production	Agents de maîtrise et chefs d'équipe	271,275 272,276 277
	P E R S O N N E L  Q U A L I F I E	Services  administratifs	Comptables et caissiers	606 607
			Secrétaires, sténo- dactylo, dactylogr.	603 609 610
			Autres personnels administratifs qualifiés	604 605 608 611
		Services commerc.	Empl.Commerce qual.	707
		Service du traite- ment de l'informa- tion	Opérateurs sur ordi- nateur. Mécanogr.calc.perfor.	622 623 624
		Services généraux	Maçons, peintres; menuisiers	320 340 463
			Conducteurs de poids lours et véhic.autom.	513 514
		Magasinage et manutention	Ouvriers et employés qualifiés du magasi- nage et de la manut.	500 501 503
		Entretien	Ouvr.mécaniciens,élec- triciens, électroméca- niciens d'entretien	384 391 394
		de la production	Ouvr.qualifiés de la forge, de la soudure, de la serrurerie, de la chaudronnerie et de la charpente de fer	360 361 362 370 371 372
			Ouvriers qualifiés de la mécanique générale	380 382
			Ouvriers qualifiés sur machines-outils	383 387 388 389
			Autres ouvriers qualifiés	341 470
	PERSONNEL SANS QUA- LIFICATION PARTICU- LIERE	Services administratifs	Empl.bureau non qualifiés	612
		Services généraux	Personnel de service et de gardiennage	951 952
		Magasinage, manutention	Manutentionnaires non qualifiés	502
		Exécution de la production	Ouvriers spécialisés Manoeuvres	471 472

LA CLASSIFICATION FONDAMENTALE DES EMPLOIS EN VINGT POSTES

Désignation des postes	Affectations à titre principal	Ventilation exhaustive
1. Cadres administratifs supérieurs.	100, 600 à 602	100, 110 à 112, 600 à 602
2. Employés qualifiés des services administratifs	603 à 611	603 à 611, 743
3. Employés de bureau sans qualification particulière	612	612
4. Cadres supérieurs et techniques des serv.commerciaux	700, 701	700, 701, 730
5. Agents techniques des services commerciaux	702 à 704	702 à 705, 731
6. Employés qualifiés des services commerciaux	707	706 à 708, 732
7. Cadres supérieurs et techniques du traitement de l'information	620, 621	207, 217, 227, 620, 621
8. Employés qualifiés du traitement de l'information	622 à 624	622 à 624
9. Personnel des services à caractère social	900, 904, 941	742, 900, 904, 910, 940 à 943
10. Personnel qualifié des services généraux	320, 340, 463, 513, 514	007, 320 à 324, 340, 423, 463, 510 à 514, 519
11. Personnel non qualifié des services généraux	951, 952	933, 950 à 953

Si l'on crée en outre une catégorie de "métiers divers", on obtient un découpage fondamental en vingt postes, susceptible de regrouper toutes les rubriques de la nomenclature. On sait que les rubriques rarement utilisées dans les déclarations correspondent d'une façon générale à des spécialités étrangères au processus de production mécanique : elles seront donc, pour la plupart, rattachées au poste "métiers divers". Cependant, chaque fois qu'un rapprochement avec un poste regroupé apparaît significatif, il sera procédé à une classification selon un critère d'assimilation. C'est ainsi notamment que tous les ingénieurs et cadres techniques supérieurs pourront être répartis entre les postes "ingénieurs et cadres techniques" d'une part et "cadres supérieurs et techniques du traitement de l'information" d'autre part ; que tous les ouvriers qualifiés pourront être répartis entre les postes "d'ouvriers qualifiés de la production et de "personnel qualifié des services généraux". On remarquera que le poids représenté par les rubriques marginales est extrêmement faible et que les discussions relatives à leur affectation présentent un intérêt d'autant plus négligeable que leur incidence structurelle est peu marquée.

Nous nous trouvons dès lors en présence :

- d'une grille détaillée fournie par la nomenclature des emplois et présentant une liste d'environ 200 rubriques plus ou moins proches les unes des autres ;

- d'une classification fondamentale en vingt postes, obtenue par regroupement des rubriques existantes, et reposant sur une combinaison possible des deux critères de la qualification requise et du service d'appartenance.

Le passage des 200 rubriques aux vingt postes regroupés est toujours possible (par construction même) mais les degrés successifs du système d'emboîtement, les modalités du rapprochement, ne peuvent être fixés sans une marge d'arbitraire. Nous retrouvons en effet à cet autre niveau la combinaison spécialité-qualification-service qui sert à définir le

12. Ingénieurs et cadres techniques	203, 206, 223, 226	200 à 206, 208, 210 à 216, 218, 220 à 226, 228
13. Techniciens	242, 250, 260, 261, 280, 284, 288.	230 à 261, 270, 280 à 288
14. Personnel d'encadrement de la production	271, 272, 275, 276, 277	271 à 277
15. Personnel qualifié du magasin. et manutention	500, 501, 503	500, 501, 503
16. Personnel non qualifié du magasin. et manutention	502	502
17. Ouvriers qualifiés d'entretien	384, 391, 394	384, 391, 394
18. Ouvriers qualifiés de la production	360 à 362, 370 à 372, 380, 382, 383, 387 à 389, 341, 470	300 à 313, 330, 331, 341 à 383, 385 à 390, 392; 393, 395 à 422, 424 à 462, 464 à 468, 470, 850, 851.
19. Ouvriers sans qualification particulière	471, 472	471, 472
20. Métiers divers	000 à 006, 008 à 030, 515 à 518, 710 à 726, 740, 741, 800 à 843, 852 à 861, 901 à 903, 905 à 909, 920, 921, 930 à 932, 960 à 999	

contenu de chaque rubrique et aucune hiérarchie ne semble plus significative qu'une autre, a priori.

## 2) Matrice des distances qualitatives

Notre objectif consiste à comparer des structures d'emploi entre elles et d'une façon plus générale à rechercher s'il n'existe pas, dans l'ensemble des modes organisationnels, des groupes d'établissements ou d'entreprises présentant certaines similitudes? Or, il ne semble pas que deux structures puissent être rigoureusement identiques (ou du moins, cela constitue un événement extrêmement rare dont la mise en évidence est d'un intérêt limité), et en sens inverse, toutes les structures offrent des éléments de similitude (condition nécessaire pour procéder à une comparaison). Le problème général reste donc entièrement posé : que signifie comparer des structures entre-elles ?

A chaque rubrique correspond, pour un établissement donné, un effectif positif ou nul, mesuré en valeur absolue ou en pourcentage. La comparaison de ces effectifs permet de calculer une distance entre deux établissements et d'une façon générale, de calculer tous les termes de la matrice des distances séparant les établissements deux à deux, point de départ de la recherche des groupes structurels éventuels. Mais s'il est clair que deux établissements séparés par une faible distance sont susceptibles de présenter des structures relativement voisines, peut-on considérer que la réciproque est nécessairement vérifiée ? Ou, en supposant que la "distance" de l'établissement B à l'établissement A est supérieure à la "distance" de l'établissement C à l'établissement A, est-on certain que les structures de A et C sont nécessairement plus voisines que les structures de A et B ? La mesure d'une telle distance est directement liée au choix des rubriques de la nomenclature et donc aux critères sous-jacents ayant servi à son élaboration. On peut seulement affirmer que par rapport au système de représentation de la réalité économique que nous avons adopté, l'instrument de mesure mis en oeuvre

fait apparaître l'établissement C plus proche de A que B. Mais, si l'on modifie le système de représentation, les résultats sont susceptibles de différer sensiblement (B peut notamment devenir plus proche de A que C). De plus, pour un système de représentation donné, l'appréciation des distances est purement quantitative et néglige entièrement les proximités relatives pouvant exister entre le contenu de deux emplois différenciés : la substituabilité éventuelle de deux rubriques constitue une excellente illustration de cet aspect du problème.

Prenons un exemple simplifié :

Rubriques	A	B	C
1	20 %	30 %	25 %
2	20 %	10 %	25 %
3	40 %	30 %	35 %
4	20 %	30 %	15 %
Total	100 %	100 %	100 %

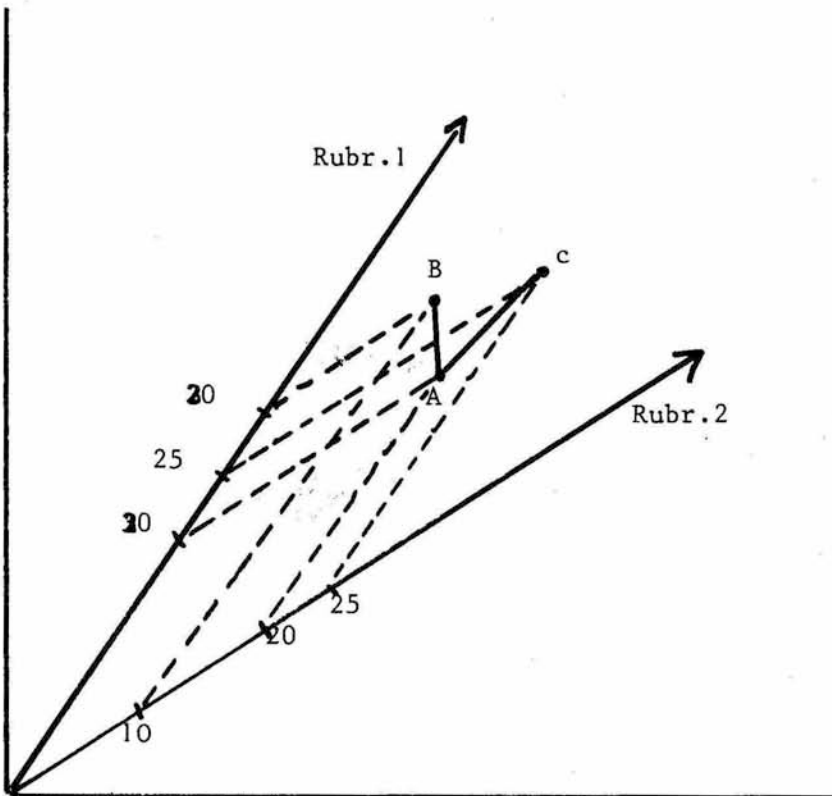
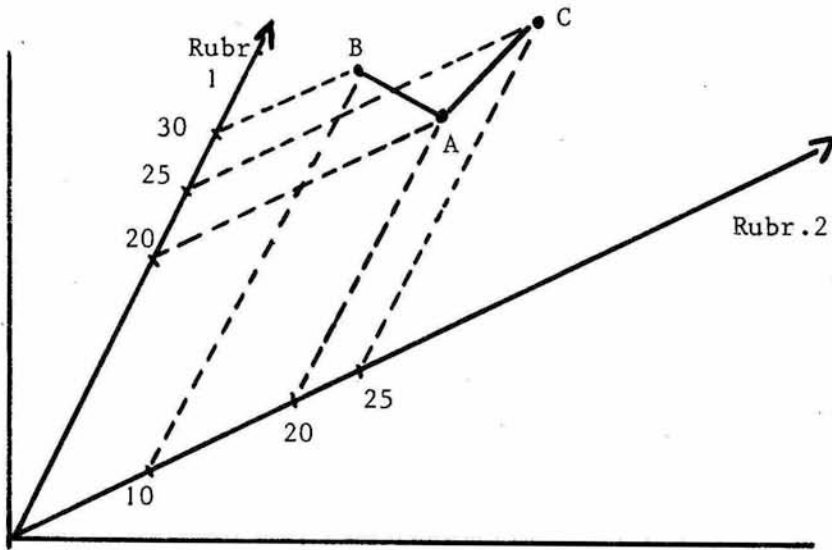
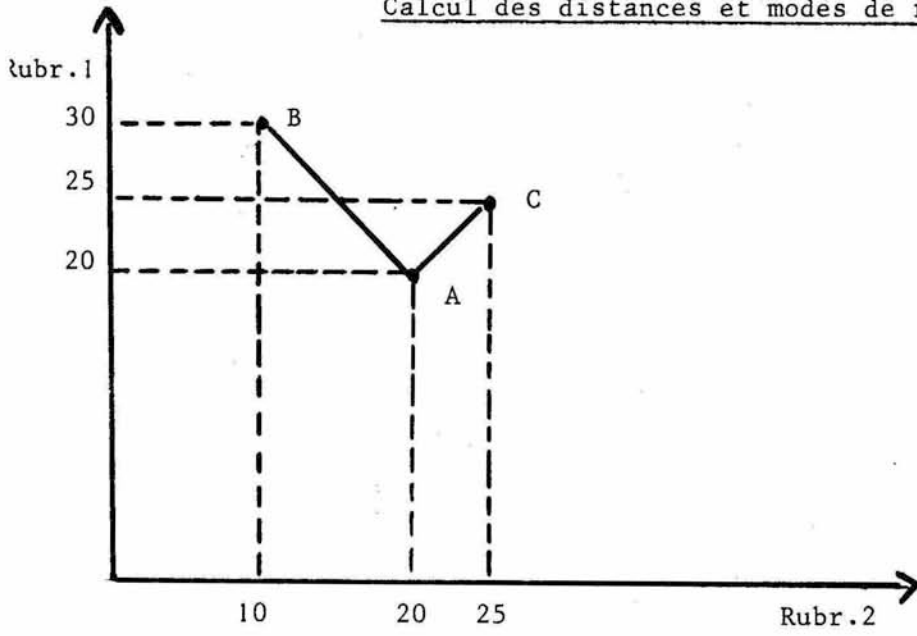
Le calcul direct exprime une distance B à A deux fois plus forte que la distance C à A. Si l'on regroupe les rubriques 1 et 2 et les rubriques 3 et 4, la distance B à A devient nulle, tandis que la distance C à A augmente sensiblement. Que conclure ? En supposant que les rubriques 1 et 2 d'une part, 3 et 4 d'autre part sont étroitement liées, il semble satisfaisant d'admettre que B est plus proche de A que C. Est-il pour autant satisfaisant d'admettre que la distance (A, B) est nulle ? Des considérations de bon sens amènent à penser que la distance (A, B) est d'autant plus faible, et corrélativement la distance (A, C) est d'autant plus forte, que le contenu des rubriques 1 et 2 et des rubriques 3 et 4 seront voisins. A la notion de distance quantitative vient s'ajouter une notion de distance qualitative qui la pondère.

Le système traditionnel de mesure de la distance entre deux établissements est le suivant. Si l'on retient  $n$  rubriques significatives, chaque établissement peut être représenté par un point dans l'espace à  $n$  dimensions.

La distance séparant deux points représentatifs mesure de façon synthétique l'écart structurel entre les deux établissements correspondants. Le carré de la distance globale est égal à la somme des carrés des distances terme à terme, le calcul pouvant être pondéré en fonction de la dispersion observée sur chacune des variables significatives : une telle démarche conduit, comme nous l'avons vu, à négliger les notions de proximité et de distorsion qualitatives, qui constituent semble-t-il, un complément d'analyse nécessaire à l'approche purement quantitative. Il s'agit pour nous, en reprenant l'exemple précédent, de bien mettre en évidence les proximités relatives existant entre les établissements A et B si l'on considère toujours que les rubriques 1 et 2 d'une part, et 3 et 4 d'autre part, recouvrent des catégories d'emplois au contenu assez peu différencié. La représentation en coordonnées orthonormées sépare nettement les points A et B et rapproche les points A et C. En revanche, la référence à un système d'axes non orthogonaux, doit permettre, par le rapprochement des axes 1 et 2 et des axes 3 et 4 d'adapter le mode de représentation à la description du phénomène économique réel (voir graphiques joints).

Considérons à nouveau la classification fondamentale en 20 postes. Il semble satisfaisant d'admettre par exemple que les cadres administratifs supérieurs sont plus proches qualitativement des ingénieurs et cadres techniques que des ouvriers qualifiés de la production. Pourquoi ? Parce que, d'une façon générale, la durée d'acquisition des qualités requises pour tenir un emploi de cadre administratif supérieur est plus courte dans le premier cas que dans le deuxième. En pratique, le problème de la détermination du temps de passage nécessaire est délicat, et de plus la relation liant deux postes entre eux

Calcul des distances et modes de représentation graphique





n'est pas forcément biunivoque : l'importance et la complexité des qualités complémentaires à acquérir pour passer d'un poste A à un poste B, ne sont généralement pas identiques à celles nécessaires pour passer du poste B au poste A (pour autant que l'idée même du passage ait un sens). Il semblerait préférable de substituer à un tel critère difficile à manier, un autre critère ou une combinaison de critères plus aisément saisissables. Les cadres administratifs supérieurs et les ingénieurs et cadres techniques se différencient par la spécialité exercée et par leur service d'appartenance ; ils se rapprochent par le niveau de formation requis pour tenir leurs emplois respectifs. Les cadres administratifs supérieurs et les ouvriers qualifiés de la production se différencient par le niveau de qualification, la spécialité exercée et le service d'appartenance. Il semblerait donc nécessaire d'établir une triple échelle des distances, construite autour des trois axes correspondant à la qualification, le service et la spécialité.

Nous admettons, à ce niveau d'analyse, que l'échelle adoptée, visant à coder des différences qualitatives, est purement arbitraire. En pratique, le but de l'opération consiste essentiellement à hiérarchiser des positions relatives de chaque poste par rapport à l'ensemble des autres, afin de moduler les résultats obtenus par application brute de calcul des distances quantitatives, en fonction du contenu réel des catégories d'emploi traitées : quelle que soit l'imprécision de l'échelle adoptée et les modalités de sa détermination, l'élément correctif introduit ne pourra qu'améliorer la qualité des résultats définitifs, même si la procédure utilisée perd une partie de sa rigueur première. A titre d'illustration, si nous adoptons une échelle de 1 à 5 sur l'axe des qualifications, une échelle de 1 à 5 sur l'axe des services et une échelle de 1 à 4 sur l'axe des

spécialités (1), nous pouvons construire la matrice jointe des distances qualitatives séparant les catégories d'emploi deux à deux. Les cadres administratifs supérieurs apparaissent ainsi relativement proches des cadres supérieurs et techniques du traitement de l'information, et dans une moindre mesure des cadres supérieurs et techniques des services commerciaux et des ingénieurs et cadres techniques ; ils apparaissent en revanche très éloignés (distance maximale) des ouvriers sans qualification particulière rattachés à la production ou au service du magasinage et de la manutention.

---

(1) La distance maximale totale entre deux postes sera dès lors, si l'on considère que les effets sont multiplicatifs, égale à 100. L'écart qualitatif séparant les cadres administratifs supérieurs des ouvriers sans qualification particulière pourra ainsi s'élever à 100 ( 5 x 5 x 4) Entre les cadres administratifs supérieurs et les ingénieurs et cadres techniques, l'application de la méthode donne un écart qualitatif de 12 ( 1 x 3 x 4)

Echelles hypothétiques utilisées dans le cadre de l'établissement  
des distances qualitatives

Qualifications	1	2	3	4	5	6	7	8
1. Ingénieurs et cadres	1	2	3	4	4	5	5	5
2. Techniciens et agents tech.	2	1	2	3	3	4	4	5
3. Personnel d'encadrement	3	2	1	2	2	3	3	5
4. Employés qualifiés	4	3	2	1	2	2	3	5
5. Ouvriers qualifiés	4	3	2	2	1	3	2	5
6. Employés non qualifiés	5	4	3	2	3	1	2	5
7. Ouvriers non qualifiés	5	4	3	3	2	2	1	5
8. Métiers divers	5	5	5	5	5	5	5	1

Services	1	2	3	4	5	6	7	8	9	10	11
1. Serv. admin.	1	3	2	4	3	3	5	5	5	5	5
2. Serv. commerciaux	3	1	3	5	4	4	5	3	5	5	5
3. Trait. inform.	2	3	1	5	5	2	5	4	5	5	5
4. Serv. à car. soc.	4	5	5	1	5	5	5	5	5	5	5
5. Serv. généraux	3	4	5	5	1	5	5	4	5	5	5
6. Org. de la prod.	3	4	2	5	5	1	2	4	3	2	5
7. Encadr. de prod.	5	5	5	5	5	2	1	3	4	2	5
8. Mag. et manut.	5	3	4	5	4	4	3	1	4	3	5
9. Entretien	5	5	5	5	5	3	4	4	1	3	5
10. Production	5	5	5	5	5	2	2	3	3	1	5
11. Métiers divers	5	5	5	5	5	5	5	5	5	5	1

Spécialité : Au niveau d'agrégation fourni par le regroupement en vingt postes, nous avons considéré que toutes les spécialités différaient fortement. La distance qualitative sur la spécialité a dans tous les cas été codée 4 (valeur adoptée maximale).

Matrice des distances qualitatives  
(regroupement en vingt postes)

Désignation des emplois	1	2	3	4	5	6
1. Cadres administratifs supérieurs	1	16	20	12	24	48
2. Employés qualifiés des serv.administrat	16	1	8	48	36	12
3. Employés de bureau sans qual.particul.	20	8	1	60	48	24
4. Cadres supérieurs et techn.des serv.Com	12	48	60	1	8	16
5. Agents techniques des serv.commerciaux	24	36	48	8	1	12
6. Employés qualifiés des serv.commerc.	48	12	24	16	12	1
7. Cadres sup. et tech. du trait.inform.	8	32	40	12	24	48
8. Empl.qualifiés du traitement de l'inf.	32	8	16	48	36	12
9. Personnels des serv.à caractère social	16	64	80	20	40	80
10. Pers.qualifiés des services généraux	48	24	36	64	48	32
11. Pers.non qualifiés des serv.Généraux	60	24	12	80	64	32
12. Ingénieurs et cadres techniques	12	48	60	16	32	64
13. Techniciens	24	36	48	32	16	48
14. Personnels d'encadr.de la production	60	40	60	60	40	40
15. Person.qualifiés du magasin.et manut.	80	20	40	48	36	12
16. Person.non qual.du magasin. et manut.	100	60	40	60	48	36
17. Ouvriers qualifiés d'entretien	80	40	60	80	60	40
18. Ouvriers qualifiés de la production	80	40	60	80	60	40
19. Ouvriers sans qualification partic.	100	60	40	100	80	60
20. Métiers divers	100	100	100	100	100	100

\* Les indices expriment la distance qualitative entre les postes (degré de différenciation relative entre le contenu des emplois désignés). Ils ont été établis à partir d'hypothèses spécifiques jouant sur les trois critères de la qualification, du service d'appartenance et de la spécialité exercée. Leurs valeurs s'étendent entre 1 (indice de similitude) et 100 (indice correspondant à la plus forte divergence).



## II - COMPARAISON STRUCTURELLE ET ANALYSE QUALITATIVE

### 1 - Calcul des distances

De tels résultats ne font que confirmer ce que nous savons déjà : cela revient en quelque sorte à expliciter de simples relations à bon sens. Comment introduire dans le calcul des distances entre établissements, l'intensité des distorsions qualitatives ? Il faut que l'instrument de mesure tienne compte à la fois de l'échelle quantitative (différences d'effectif portant sur des grandeurs directement comparables) et de l'échelle qualitative (pondération des résultats bruts en fonction des proximités relatives existant entre le contenu des emplois désignés). La dimension horizontale (différences terme à terme) doit être complétée par une dimension verticale (différences de contenu entre les termes). Pour un établissement, l'effectif rattaché à un poste donné ne prend toute sa signification que s'il est confronté aux effectifs de tous les autres postes occupés : la comparaison de deux structures ne deviendra elle-même opératoire que si elle porte sur des indices exprimant correctement l'ordonnement structurel réel.

Dans l'exemple simplifié développé au paragraphe précédent, les structures des établissements A, B et C ne semblent pas devoir être comparées directement. Une première opération, déjà effectuée, vise à transformer les effectifs absolus en pourcentages, en rapportant les effectifs de chaque rubrique à l'effectif total de l'établissement. Une deuxième opération apparaît en outre nécessaire. La rubrique 1 par exemple regroupe 20 % des effectifs de A, 30 % des effectifs de B et 25 % des effectifs de C : la comparaison directe sur ces trois pourcentages apporte certes un des éléments d'appréciation de l'écart structurel entre les trois établissements, mais nous avons vu que les différences brutes ne pouvaient être appréciées significativement sans tenir compte des effectifs correspondants de la rubrique 2 (20 %, 10 %, 25 %). Il en va de même des comparaisons portant sur la rubrique 2 (interaction avec 1), sur la rubrique 3 (interaction avec 4) et sur la rubrique 4 (interaction avec 3). De sorte que, si  $\alpha$  mesure l'intensité de la liaison rubrique 1 - rubrique 2 et  $\beta$  l'intensité de la liaison rubrique 3 - rubrique 4, le phénomène étudié pourra être mis en évidence en procédant aux comparaisons successives des termes suivants :

$$\left( \begin{array}{l} \text{EFF (R}_1\text{)} + \alpha \text{ EFF (R}_2\text{)} \\ \text{EFF (R}_2\text{)} + \alpha \text{ EFF (R}_1\text{)} \\ \text{EFF (R}_3\text{)} + \beta \text{ EFF (R}_4\text{)} \\ \text{EFF (R}_4\text{)} + \beta \text{ EFF (R}_3\text{)} \end{array} \right)$$

Soit, en définissant la matrice des proximités qualitatives :

Rubr.	1	2	3	4
1	1	$\alpha$	0	0
2	$\alpha$	1	0	0
3	0	0	1	$\beta$
4	0	0	$\beta$	1

les comparaisons devront porter sur les quatre termes du vecteur défini comme le produit du vecteur structure par la matrice des proximités qualitatives.

A titre d'illustration, nous donnons les résultats du calcul des distances (A, B) et (A, C) pour différentes valeurs de  $\alpha$  et  $\beta$

$\alpha$	$\beta$	d(A,B)	d(A,C)
0	0	20	10
0,5	0,5	10	15
0,8	0,8	4	18
1	1	0	20

D'une façon générale, la matrice des proximités est directement dérivée de la matrice des distances (deux postes très "distants" doivent être caractérisés par un indice de proximité très faible et inversement). Il nous a été ainsi possible, en utilisant une formule de transformation (1) de chacun des termes de la matrice des distances précédemment établie, de construire une matrice des proximités qualitatives liant deux à deux les vingt postes de la classification fondamentale. Il est bien clair que les chiffres présentés n'ont aucune portée définitive, dans la mesure où les hypothèses sous-jacentes (échelles des distances qualitatives) sont toujours susceptibles de

---

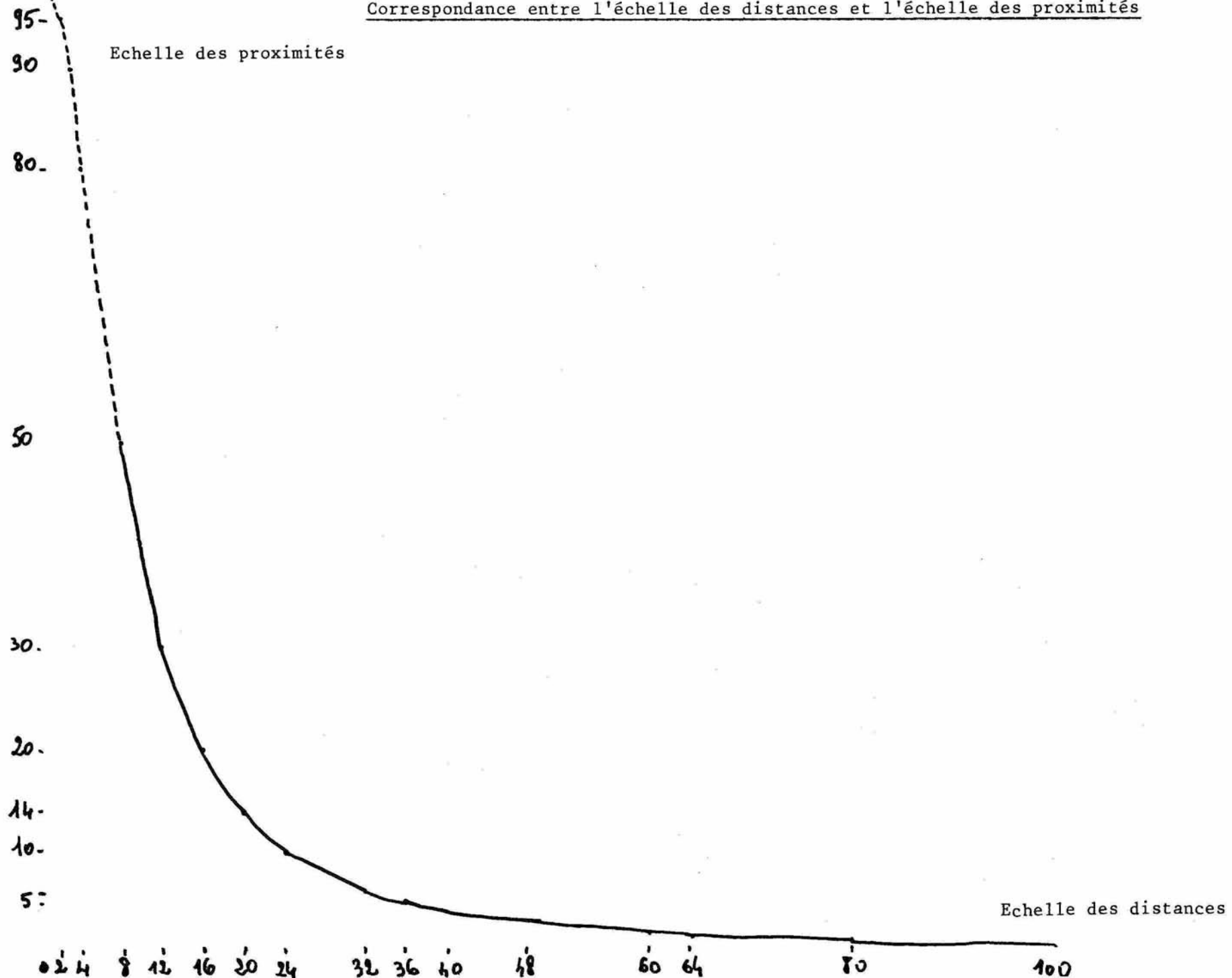
(1) En appelant X l'indice de la distance qualitative entre deux catégories d'emploi, les termes de la matrice des proximités qualitatives ont été calculés par application de la formule :

$$\frac{100}{1 + \left(\frac{X}{8}\right)^2}$$

Les valeurs obtenues ont été ramenées à une base 100 (indice de similitude) et arrondies à l'unité la plus proche. L'indice mesurant la proximité qualitative entre les cadres administratifs supérieurs et les ouvriers sans qualification particulière par exemple est égal à Q (valeur minimale). Entre les cadres administratifs supérieurs et les ingénieurs et cadres techniques, sa valeur est de 30 (comme entre cadres administratifs supérieurs et cadres supérieurs et techniques des services commerciaux).



Correspondance entre l'échelle des distances et l'échelle des proximités



Matrice des proximités qualitatives  
(regroupement en 20 postes)

Désignation des emplois	1	2	3	4	5	6
1. Cadres administratifs supérieurs	100	20	14	30	10	3
2. Employés qualifiés des serv.administr	20	100	50	3	5	30
3. Employés de bureau sans qual.partic.	14	50	100	2	3	10
4. Cadres sup.et tech. des serv. Comm.	30	3	2	100	50	20
5. Agents techniques des serv. Comm.	10	5	3	50	100	30
6. Employés qualifiés des serv.Comm.	3	30	10	20	30	100
7. Cadres sup.et tech. trait. de l'inf.	50	6	4	30	10	3
8. Empl.Qualifiés du traitement de l'inf	6	50	20	3	5	30
9. Personnels des serv.à caract. social	20	1	1	14	4	1
10. Pers .qualifiés des services généraux	3	10	5	1	3	6
11. Pers.non qualifiés des serv.généraux	2	10	30	1	1	6
12.Ingénieurs et cadres techniques	30	3	2	20	6	1
13. Techniciens	10	5	3	6	20	3
14. Personnels d'encadrement de la prod.	2	4	2	2	4	4
15. Pers. Qualifiés du magas. et manut.	1	14	4	3	5	30
16. Pers.non qual. du magas. et manut.	0	2	4	2	3	5
17. Ouvriers qualifiés d'entretien	1	4	2	1	2	4
18. Ouvriers qualifiés de la production	1	4	2	1	2	4
19. Ouvriers sans qualif. particulière	0	2	4	0	1	2
20. Métiers divers	0	0	0	0	0	0



de modifications éventuelles : ils visent essentiellement à fixer les idées et à illustrer les étapes du raisonnement appelé à constituer le support de notre recherche. Les comparaisons structurelles devront reposer sur une comparaison deux à deux des vecteurs structure quantitative-qualitative : ces vecteurs étant définis pour chaque établissement comme le produit du vecteur structure quantitative par la matrice des proximités qualitatives.

En d'autres termes, les éléments sur lesquels devront porter les comparaisons directes apparaissent comme des combinaisons linéaires de l'ensemble des éléments constitutifs de la structure brute des établissements. Les grandeurs ainsi définies ne présentent pas en propre une signification économique concrète : elles fournissent seulement des données intermédiaires nécessaires à la mesure des distorsions structurelles existant entre les établissements. Les écarts entre les termes bruts seront ou atténués ou amplifiés en fonction du contexte organisationnel réel, créant un effet de distanciation quantitative-qualitative soit plus important, soit moins important que dans le cas de la mesure purement quantitative : plus important si les divergences sont profondes, moins important si elles sont superficielles. On doit insister sur le caractère fondamental des hypothèses utilisées : un bouleversement de ces hypothèses est susceptible en effet de modifier sensiblement la hiérarchie des établissements entre eux. Cela signifie d'abord que leur élaboration mérite certainement une plus longue recherche que celle que nous avons entreprise jusqu'à présent. Cela signifie également que l'instrument mis en oeuvre offre une grande souplesse d'adaptation en permettant de privilégier certains critères (qualification par exemple) par rapport aux autres.

Au stade actuel, nous nous efforçons principalement de développer l'analyse méthodologique en retenant la possibilité de jouer ultérieurement sur le choix des relations qualitatives. A chaque établissement est associé un vecteur structure quantitative-qualitative.

Comparer deux établissements revient à comparer deux vecteurs : par définition, le carré de la distance quantitative-qualitative séparant deux établissements est égal à la somme des carrés des écarts entre les termes correspondants des vecteurs structure quantitative-qualitative. On peut ajouter en outre, si l'on veut prolonger l'analyse, qu'une différence portant sur un cadre administratif supérieur par exemple ne présente ni la même signification, ni la même importance qu'une différence portant sur un ouvrier sans qualification particulière. D'une part parce que la dispersion sur la variable "ouvriers sans qualification particulière" est sensiblement plus forte que la dispersion sur la variable "cadres administratifs supérieurs". D'autre part, parce que les niveaux de qualification requis pour tenir ces deux catégories d'emploi ne sont pas les mêmes. Il semblerait donc éventuellement souhaitable d'introduire une double pondération de la mesure des distances quantitative-qualitatives :

- une pondération quantitative, qui ramènerait la mesure des distances terme à terme à une échelle plus directement comparable (il suffit pour cela de normer chacune des distances par l'écart type) ;

- une pondération qualitative, qui tendrait à accorder un poids variable à un écart donné selon la catégorie d'emploi sur laquelle il porte (l'échelle la plus simple à établir, et la plus facile à justifier, est une échelle reposant sur le niveau de qualification : on accorderait d'autant plus de poids à une différence donnée qu'elle porterait sur une catégorie d'emploi requérant une qualification élevée).

Considérons deux établissements  $i$  et  $j$ . Appelons  $X_k^i$  et  $X_k^j$  les termes de rang  $k$  des vecteurs structure quantitative-qualitative associés aux deux établissements  $i$  et  $j$  ;  $\sigma_k$  l'écart type relatif à la variable  $k$  ; et  $I_k$  l'indice de pondération qualitative des écarts observés sur  $k$ . Il découle des développements précédents que le carré de la distance entre  $i$  et  $j$  est égal à :

$$d^2(i, j) = \sum_k I_k \frac{(X_k^i - X_k^j)^2}{\sigma_k^2}$$

2 - Technique de regroupement des rubriques

Le problème du regroupement des postes d'une nomenclature s'inscrit dans le prolongement des développements précédents. En règle générale, la distance entre les termes regroupés est considérée nulle tandis que les autres distances sont toutes supposées égales entre elles. Il s'agit là d'une simplification quelque peu hardie de la démarche analytique qui semble devoir s'imposer. Chaque rubrique est définie par sa fréquence (critère quantitatif) et par sa distance relative aux autres rubriques (critère qualitatif). Un tel principe rend compte du degré de similitude susceptible d'exister entre le contenu de deux emplois différenciés, et il semblerait n'y avoir a priori aucune raison théorique justifiant le regroupement de deux rubriques, hormis un souci bien légitime de simplification des calculs. Si une analyse portant sur une nomenclature regroupée doit fournir des résultats sensiblement identiques à ceux qu'il serait possible de dégager en raisonnant sur l'ensemble des rubriques détaillées, il est clair que le regroupement s'impose puisqu'il permet de réduire le temps de traitement et d'économiser du temps machine. On doit donc regrouper si et si seulement :

$$\frac{\frac{d(i, j)}{d(i, k)}}{\frac{d'(i, j)}{d'(i, k)}} \neq 1 \quad \text{pour } \forall i, j, k$$

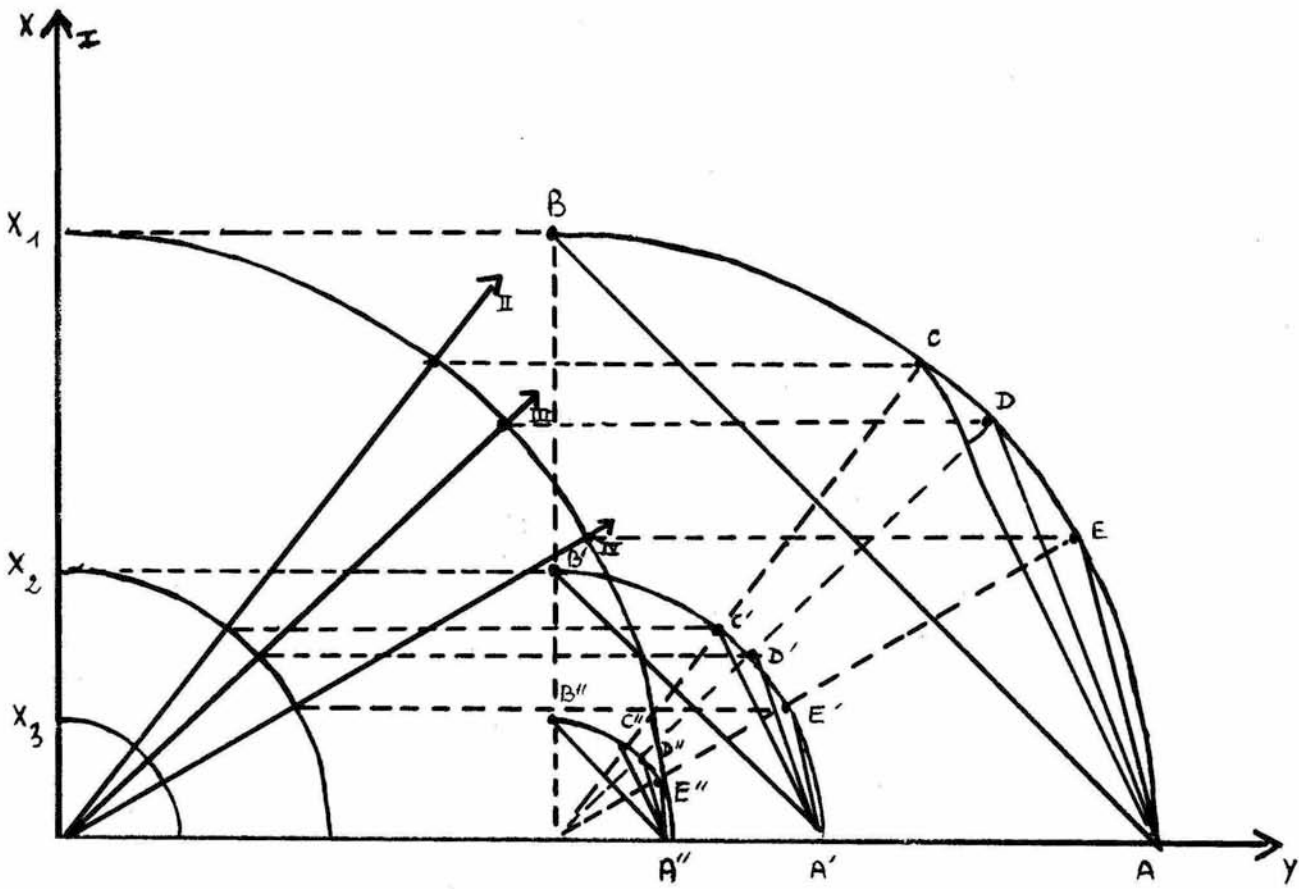
avec :

	Nomenclature détaillée	Nomenclature regroupée
Distance de l'éts i à l'éts. j	d (i, j)	d' (i, j)
Distance de l'éts i à l'éts k	d (i, k)	d' (i, k)

La résolution directe et exhaustive d'un tel problème par passages successifs en machine ne présente aucun intérêt dans la mesure où elle suppose au préalable le traitement sur nomenclature détaillée que l'on souhaiterait éviter. Deux orientations sont possibles : utiliser la démarche directe sur un échantillon réduit (une centaine d'établissements tirés au hasard) ; ou définir une méthode d'approche simplifiant la problématique générale. En pratique, les deux orientations sont étroitement liées. Le rapprochement de certaines rubriques entre elles est dépourvu de toute signification économique : il est inutile par exemple "d'essayer" le regroupement directeur d'établissement et ouvriers sans qualification particulière. En revanche, certains regroupements seront privilégiés par rapport aux autres. On retrouve à ce niveau les hypothèses implicitement adoptées dans le cadre des recherches de nomenclatures simplifiées : on considère en général, qu'une rubrique peu utilisée (par un petit nombre d'établissements ou pour un petit nombre de salariés) doit être regroupée avec une autre rubrique ; et que les regroupements doivent s'effectuer entre rubriques "voisines" afin de ne pas modifier sensiblement les écarts structurels existants.

De telles hypothèses peuvent être illustrées graphiquement. Soit une rubrique X et l'on s'interroge sur son regroupement éventuel avec une rubrique Y. Nous avons défini douze cas correspondant à trois valeurs de X ( $X_1, X_2, X_3$  avec  $X_1 > X_2 > X_3$ ) combiné à quatre indices de distance entre X et Y (I, II, III et IV avec  $dI > dII > dIII > dIV$ ).

Illustration graphique du problème du regroupement  
de deux rubriques





Nous avons :

Distance qualit. X/Y	I	II	III	IV	I	II	III	IV	I	II	III	IV
Effectif de rubr. X	X <sub>1</sub>	X <sub>1</sub>	X <sub>1</sub>	X <sub>1</sub>	X <sub>2</sub>	X <sub>2</sub>	X <sub>2</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>3</sub>	X <sub>3</sub>	X <sub>3</sub>
P <sup>t</sup> représ. <u>sans</u> regroup	B	C	D	E	B'	C'	D'	E'	B''	C''	D''	E''
P <sup>t</sup> représ. <u>avec</u> regroup	A	A	A	A	A'	A'	A'	A'	A''	A''	A''	A''
Perte d'information lié au regroupement.	$\overline{AB}$	$\overline{AC}$	$\overline{AD}$	$\overline{AE}$	$\overline{A'B}$	$\overline{A'C}$	$\overline{A'D}$	$\overline{A'E}$	$\overline{A''B''}$	$\overline{A''C''}$	$\overline{A''D''}$	$\overline{A''E''}$

Le regroupement de la rubrique X avec la rubrique Y apparaît dès lors d'autant plus abusif que :

- l'effectif de la rubrique X est important

$(\overline{AB} > \overline{A'B'} > \overline{A''B''}, \dots, \overline{AE} > \overline{A'E'} > \overline{A''E''})$

- la distance qualitative entre X et Y est importante  $(\overline{AB} > \overline{AC} > \overline{AD} > \overline{AE}, \dots, \overline{A''B''} > \overline{A''C''} > \overline{A''D''} > \overline{A''E''})$

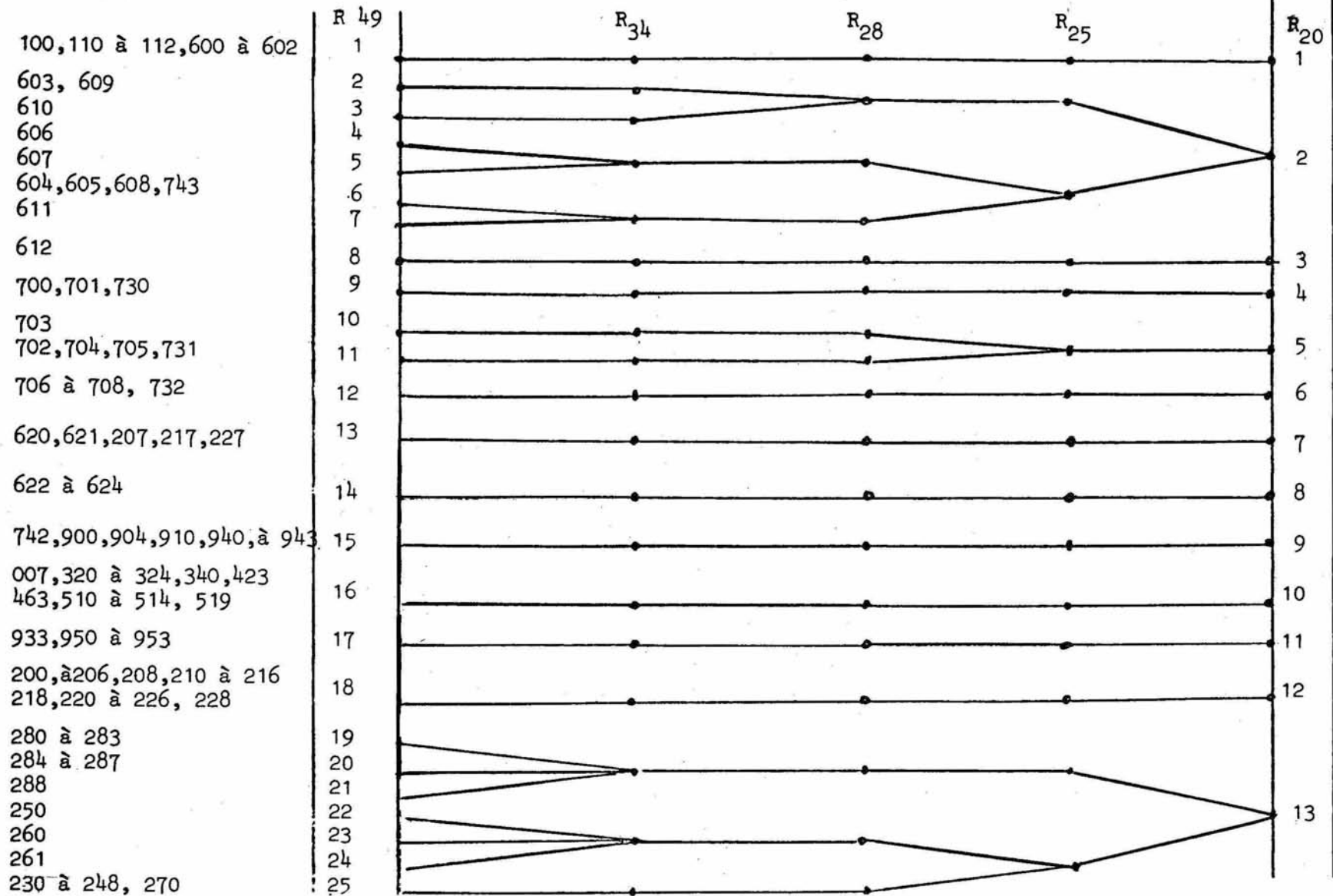
Ou, en inversant le problème, l'incitation au regroupement sera d'autant plus grande que l'effectif de la rubrique à regrouper sera faible et que les deux rubriques appelées à être regroupées désigneront des catégories d'emploi au contenu voisin. Ainsi, les rubriques rarement (et a fortiori jamais) utilisées pourront être regroupées soit entre-elles, soit avec une autre rubrique plus fréquente, qualitativement proche. D'une façon générale, le regroupement de deux rubriques pourra être "accepté" s'il satisfait à certaines conditions de proximité (critère qualitatif) et d'effectif (critère quantitatif). Ou, sans une présentation différente, on peut considérer que les relations de proximité définissent les modalités d'un regroupement éventuel, alors que les effectifs des rubriques en précisent les conditions d'acceptation.

En pratique, une telle démarche, si elle présente l'avantage de se prêter au calcul systématique, n'offre en elle-même rien de profondément neuf par rapport aux recherches empiriques : ses résultats apparaissent en grande partie comme la conséquence des hypothèses adoptées (échelle des distances qualitatives) lesquelles découlent directement des études empiriques préalables. Nous disposons dans un premier temps d'une liste d'emplois détaillée (296 rubriques) et d'un regroupement significatif en 28 postes réunissant entre elles les rubriques originellement séparées par une faible distance qualitative. Une rubrique donnée ne sera insérée au groupe qui la coiffe (critère qualitatif) que si son effectif est lui-même en règle générale très faible (critère quantitatif). En fixant par exemple le niveau d'acceptation à 1 %, nous avons pu obtenir une classification des emplois en 49 postes, composés soit d'une seule rubrique lorsque son effectif dépassait en moyenne 1 % de l'effectif total (rejet du regroupement) , soit de plusieurs rubriques lorsque leur effectif propre représentait en moyenne moins de 1 % de l'effectif total (acceptation du regroupement). En fixant le niveau d'acceptation à 3 % on obtient une nouvelle classification en 34 postes. Les séquences de calcul sont déterminées par les hypothèses successives relatives aux 2 critères retenus. Ainsi, le choix des hypothèses indiquées sur le tableau ci-après, conduit à tracer un système d'emboîtement décrivant en cinq étapes (les conditions, très forte proximité - effectif quelconque et forte proximité - effectif très faible, donnant des résultats équivalents) le passage de la nomenclature détaillée à la classification fondamentale en vingt postes :

Hypothèses	NB de postes
Très forte proximité très faible effectif	49 postes
Très forte proximité Effectif faible	34 postes
Très forte proximité Effectif quelconque	28 postes
Forte proximité effectif très faible	28 postes
Forte proximité effectif faible	25 postes
Forte proximité Effectif quelconque	20 postes

A l'issue du premier regroupement (49 postes) douze des vingt postes de la classification fondamentale sont déjà définis. En fait, le système d'emboîtement (voir l'arbre joint, p. ) affecte principalement, au-delà du regroupement en 49 postes, trois catégories d'emploi : les employés qualifiés des services administratifs, les techniciens et les ouvriers qualifiés de la production. Ainsi le choix d'un regroupement donné (20, 25, 28, 34 et 49 postes) peut être dicté par des considérations purement théoriques : si l'on accepte par exemple de négliger totalement les distorsions structurelles liées à la ventilation des employés administratifs qualifiés, des techniciens et des ouvriers qualifiés de la production, la classification des emplois en vingt postes apparaît parfaitement opérante, et elle n'est opérante que si l'on accepte cette hypothèse de travail. Et, d'une façon générale, il est possible d'utiliser l'un quelconque des cadres d'analyse proposés à condition d'en préciser les limites, de souligner les catégories d'emploi sur lesquelles les comparaisons risquent d'être écrasées.

Système d'emboitage des rubriques reposant sur un double critère de proximité et d'effectif



271, 273, 275  
272, 274, 276  
277

500  
501, 503

502

384  
391, 394

361  
360, 362  
370

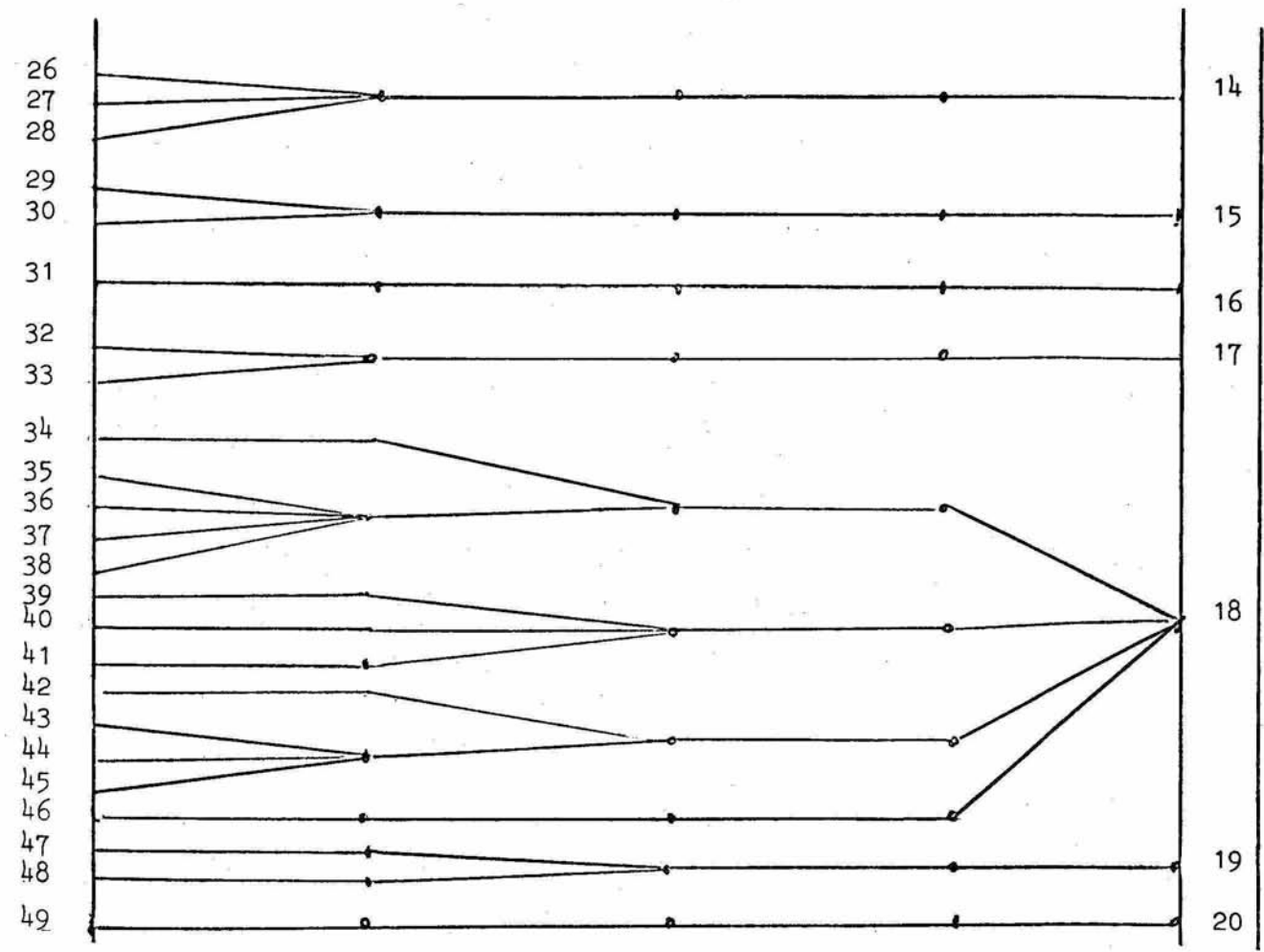
371  
372  
380  
382  
381, 385, 386

387  
383  
388  
389

Voir liste détaillée(1)

471  
472

Voir liste détaillée(2)



Liste détaillée (1) : 300 à 313, 330, 331, 341 à 352, 363, 390, 392, 393, 395 à 422  
424 à 462, 464 à 468, 470, 850, 851

Liste détaillée (2) 000 à 006, 008 à 030, 515 à 518, 710 à 726, 740, 741, 800 à 843  
852 à 861, 901 à 903, 905 à 909, 920, 921, 930 à 932, 960 à 999

*C H A P I T R E   I I I*

---

*RECHERCHE THEORIQUE POUR LA CONSTRUCTION D'UN OUTIL  
DE COMPARAISON*

## I - LA STRUCTURE DES EMPLOIS

La répartition de son personnel salarié entre les rubriques fournies par une nomenclature, définit communément la structure d'une entreprise ou d'un établissement, dont le contenu apparaît dès lors largement tributaire du mode de repérage et de représentation des informations collectées. Deux critiques, ou plutôt, deux interrogations peuvent être formulées, l'une concernant la qualité intrinsèque de la classification des activités professionnelles, c'est-à-dire sa cohérence interne et la pertinence de ses critères par rapport aux objectifs généraux poursuivis, l'autre concernant la valeur extrinsèque de l'instrument mis en oeuvre, c'est-à-dire son aptitude à décrire une situation organisationnelle réelle et à constituer un support d'analyse adapté à l'étude comparative des structures d'emploi.

Le premier point se trouve développé à travers une littérature abondante, mettant l'accent sur les principes d'homogénéité et d'exclusion qui guident la logique de la procédure, insistant sur la nécessité d'une adéquation entre finalité de l'enquête et modalité de classification des emplois, proposant enfin un aménagement ou une réforme des nomenclatures d'activités individuelles. Mais là n'est pas notre propos. Il n'est pas possible, en effet, de modifier une information déjà existante dont les défauts éventuels s'imposent à l'utilisateur. Celui-ci doit en revanche s'efforcer de construire, à partir des données disponibles, un instrument d'analyse adapté à ses propres problèmes.

La deuxième interrogation devient alors essentielle, et son intérêt dépasse largement la simple critique du choix des rubriques. C'est le principe même du clivage des emplois qui est mis en cause, leur repérage sous la forme d'une liste d'appellations où les conditions d'enregistrement des données cachent artificiellement

l'équilibre consacré par le concept d'unité organisationnelle, risquent d'amplifier ou au contraire de réduire les distorsions structurelles observées si leur mesure est fondée uniquement sur une comparaison terme à terme des effectifs répertoriés. Une dimension manque, celle qui permettrait de situer un emploi par rapport à l'ensemble des autres ou d'ordonner d'une façon générale les diverses rubriques entre-elles, qui autoriserait la construction d'un véritable système de représentation des structures précisant la signification réelle des informations disponibles, qui allierait en un mot la donnée quantitative initiale à son expression qualitative, support indispensable d'une approche explicative.

Les efforts entrepris dans le cadre de l'amélioration des nomenclatures présentent certes une grande utilité. Ils répondent à un besoin impératif de connaissances destinées à faciliter l'action et la recherche, ils visent à aligner le principe de collecte de l'information sur la construction d'un système logique d'interrogations successives éventuellement emboîtées, ils conduisent à cerner l'ensemble des données statistiques nécessaires à la satisfaction d'objectifs généraux préalablement fixés, autant d'éléments fortement appréciés de l'utilisateur dont les travaux dépendent étroitement de la qualité du matériau de base qui lui est fourni. Mais améliorer une nomenclature, c'est seulement proposer une nouvelle classification des emplois plus satisfaisante que la précédente. Là s'arrête son ambition. Le problème de la définition d'un instrument d'analyse approprié au caractère spécifique de l'étude reste donc entier si l'on admet que le recueil de statistiques de main d'oeuvre n'apporte pas directement la solution recherchée.

Une telle remarque permet de mieux situer la portée générale de la méthode qui va être présentée. Elle indique aussi cependant quelles en sont les contraintes : une partie des informations requises sort du cadre habituel des nomenclatures et il convient de se les procurer par ailleurs. A l'heure actuelle, les données concernant notamment la liaison entre emplois sont inexistantes et les lacunes doivent être provisoirement



complées par un jeu d'hypothèses dont le fondement et la validité appellent la discussion. Est-ce une raison suffisante pour ne pas utiliser l'instrument construit ? Certainement pas si le gain d'efficacité l'emporte sur les imperfections de mesure, certainement pas si le choix des hypothèses peut être amélioré, certainement pas enfin, si les possibilités de la méthode amènent à élaborer les informations manquantes.

#### A - Modalités de représentation structurelle

Une nomenclature constitue d'une façon générale un système de représentation simplifié d'une réalité plus complexe. Son principe d'élaboration vise à regrouper sous une même rubrique des éléments supposés homogènes, afin de concentrer l'analyse sur un nombre restreint de catégories agrégées. La procédure repose sur deux concepts fondamentaux :

- la notion de distance et de proximité, qui permet de négliger les distorsions qualitatives mineures entre les éléments regroupés sous une même rubrique ;

- la notion de représentativité qui permet de raisonner sur un échantillon limité de valeurs caractéristiques, ayant la propriété de réunir une quantité suffisante d'informations par rapport aux objectifs assignés.

Ainsi, pour construire un système cohérent de description de la structure de l'emploi, il apparaît nécessaire de combiner deux types de données complémentaires :

- . des données relatives à la distribution des emplois entre les rubriques retenues ,
- . des données relatives à la distance (ou à la proximité) des rubriques entre elles.

Cela signifie que l'effectif d'une rubrique est à rapprocher non seulement de l'effectif total, mais également de l'effectif de chacune des autres rubriques par l'intermédiaire d'une clé de passage exprimant l'intensité du lien qualitatif entre rubriques. En pratique, cette deuxième phase incombe à l'analyste, qui dispose seulement d'une information brute ventilée en classes supposées exclusives les unes des autres. Sa démarche le conduit successivement :

- à s'intéresser à quelques repères caractéristiques privilégiés (soit définis directement à partir de l'effectif de certaines rubriques dominantes, soit obtenus par agrégation de plusieurs rubriques existantes).

- à ramifier sa recherche par exploration systématique des axes d'interrelation qui explicitent la configuration des noeuds significatifs initiaux ;

- et à étendre la portée des résultats dégagés sur l'échantillon descripteur pour rendre compte du mode organisationnel dans son ensemble.

En d'autres termes, le travail de l'analyste parachève le travail d'élaboration de l'instrument de mesure, qui, en utilisant comme unique support une simple liste d'appellations, présente un caractère trop incomplet pour satisfaire aux objectifs poursuivis (analyse des structures de l'emploi). L'utilisateur se substitue en quelque sorte au producteur d'information, et cela parce que les deux concepts fondamentaux sous-jacents au système proposé (distance et représentativité) ne sont pas explicités.

Une telle entreprise, non seulement utile et souhaitable, est rendue nécessaire par l'imperfection de l'outil. Elle semble toujours possible, bien que délicate, dans le cas d'un petit nombre d'observations. En présence d'un grand nombre d'observations, elle exige la définition d'une procédure systématique visant à créer un instrument de mesure plus opérant, combinant :

- les données brutes de départ (structure traditionnelle)

- l'échelle des **distances** liant les rubriques entre elles (introduction de la dimension qualitative) ;

- les modalités de généralisation (passage de la discrétisation de l'échantillon au continu).

### 1 - Les données brutes

Il ne s'agit pas ici de proposer une nouvelle nomenclature. Nous nous trouvons dans la position de l'utilisateur d'information et le choix des critères classants (division fonctionnelle, grille des qualifications, spécialité exercée) apparaît comme une donnée que nous n'avons pas vocation à modifier.

Le seul élément d'appréciation disponible est fourni par la fréquence d'utilisation de chaque rubrique. Une nomenclature apparaît en effet comme une série d'interrogations successives appelant des réponses chiffrées (0, 1, ... n individus rattachés à une rubrique déterminée). L'analyse de ces réponses doit fournir les informations recherchées sur la structure des emplois.

En pratique, la fréquence d'utilisation de certaines rubriques est nulle ou voisine de zéro. La nomenclature est en effet un référentiel adapté à des ensembles hétérogènes. Le passage à l'étude d'une sous-population d'établissements d'un secteur d'activité donné conduit à s'interroger sur la correspondance entre le schéma conceptuel adopté et la réalité économique décrite.

Plus généralement, deux sources de distorsion peuvent exister :

- l'une liée à la spécificité du secteur étudié qui influence la distribution des emplois entre les rubriques ;
- l'autre liée à l'imperfection du système de repérage des emplois qui agit sur la qualité des réponses obtenues.

Cela explique que l'on soit amené à raisonner sur une distribution structurelle simplifiée, bâtie autour d'un nombre restreint de postes caractéristiques fondamentaux, qu'il est possible d'obtenir en regroupant les rubriques initiales. Selon cette orientation, doivent être agrégées :

- les rubriques dont la fréquence d'utilisation est faible ou négligeable (principe d'adéquation à la spécificité sectorielle) ;

- les rubriques désignant des emplois "voisins" entre lesquels les confusions risquent d'être fréquentes (principe ~~et~~ d'adéquation à la réalité économique décrite).

Une telle démarche suppose l'existence, à un niveau d'agrégation supérieur, d'un système de représentation des emplois ayant la double propriété :

- d'absorber toutes les confusions initiales possibles et de permettre par là une classification unique des emplois ;

- de conserver toutes les caractéristiques structurelles fondamentales et de permettre par là une analyse significative des emplois.

Il existerait donc une nouvelle nomenclature, plus simple et mieux adaptée au problème étudié, susceptible de coiffer la nomenclature initiale en utilisant à la fois les données qu'elle fournit et les principes sur lesquels elle repose.

A notre avis, deux remarques peuvent être formulées à l'encontre d'une telle démarche simplificatrice :

- le regroupement des rubriques s'accompagne nécessairement d'une perte d'information qu'il convient de rendre minimale : la recherche d'un équilibre est incompatible avec la résolution parfaite du problème posé ;

- la relation de proximité, indirectement utilisée lors du regroupement, est transformée en une relation d'identité : la nature des résultats dégagés est en grande partie liée au choix du seuil d'assimilation retenu.

## 2) L'échelle des distances

La déclaration de l'employeur consiste à répartir son personnel salarié entre les rubriques de la nomenclature qui lui est proposée. Il s'agit d'isoler, parmi l'ensemble des caractéristiques définissant l'emploi, celles qui ont été retenues dans le cadre de l'élaboration de la nomenclature. La concordance, l'adéquation entre le système de représentation et la réalité décrite, ne présente pas toujours un degré de précision évitant toute confusion :

- il peut exister plusieurs rubriques différentes dont la définition se rapproche du contenu réel du poste occupé ;
- à une appellation donnée peut correspondre une gamme d'emplois réels assez fortement différenciés.

On admet généralement que la distance entre deux emplois affectés à une même rubrique est plus faible que la distance entre deux emplois affectés à deux rubriques différentes. Le schéma théorique de classification s'appuie en effet :

- sur une analyse préalable des emplois et des rubriques, destinée à hiérarchiser, de façon plus ou moins implicite, les écarts entre le contenu du poste réel et la distribution des tâches susceptibles d'être rattachées à chacune des rubriques proposées ;

- sur une procédure générale de choix exclusif de la rubrique d'affectation, visant à assimiler l'emploi réel à la rubrique dont le contenu semble s'en approcher le plus.

Cela revient à :

- Substituer au mode de représentation uni-dimensionnel (nomenclature d'appellations), un mode de représentation dans l'espace, faisant intervenir une échelle complexe des distances entre rubriques ;

- Représenter la position réelle de l'emploi occupé par rapport au schéma établi ;
- Confondre le point réel avec le point théorique qui s'en rapproche le plus ;
- Substituer au mode de représentation dans l'espace, un mode de représentation unidimensionnelle.

Une telle démarche ne va pas sans une marge d'arbitraire importante. La notion de distance à laquelle elle se réfère n'est pas précisée dans la nomenclature. Elle repose, théoriquement, sur une combinaison d'échelles élémentaires établies pour chacun des critères retenus, la construction de ces échelles et les modalités de leur combinaison étant laissées à la libre appréciation du déclarant . Ainsi, deux emplois différents peuvent être selon les cas, confondus ou dissociés, pour peu que l'échelle des distances adoptée tende à les rapprocher ou à les éloigner l'un de l'autre.

De sorte que :

- les "erreurs" de déclaration peuvent être attribuées dans une large mesure à l'adoption d'une échelle des distances s'écartant de la "norme" ;
- les affectations, à une rubrique donnée, d'emplois "peu différenciés" introduisent un écrasement artificiel de l'échelle des distances réelles ;
- les affectations à des rubriques différentes, d'emplois "peu différenciés" introduisent une amplification artificielle de l'échelle des distances réelles.

Il semble cependant clair (et c'est là le seul élément de certitude disponible), que d'une façon générale, deux emplois relativement "proches" seront soit rattachés à une seule et même rubrique, soit rattachés à deux rubriques "proches" l'une de l'autre. Dans le premier cas, l'assimilation conduit à négliger les

différences existantes. Pour éviter de créer une distorsion non significative et pour atténuer dans une certaine mesure l'effet de glissement de l'échelle des distances susceptibles d'apparaître, il est nécessaire, dans le deuxième cas, de préciser le sens de la classification par un indicateur de proximité des rubriques.

### 3) Les modalités de représentation structurelle

Pour simplifier le problème, supposons que l'échelle des distances adoptée permette une représentation dans le plan des positions relatives des rubriques de la nomenclature (voir graphique I). Le schéma est construit de la façon suivante :

- la position d'une rubrique quelconque A est fixée arbitrairement;

- par rapport au point A, la position des rubriques B et C est déterminée en fonction des distances (A,B), (A,C), et (B,C).

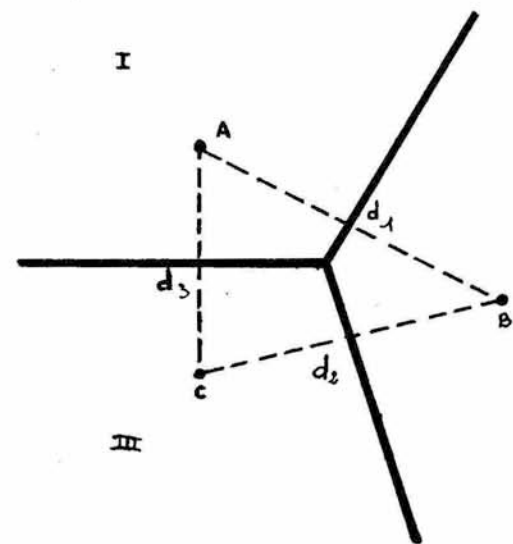
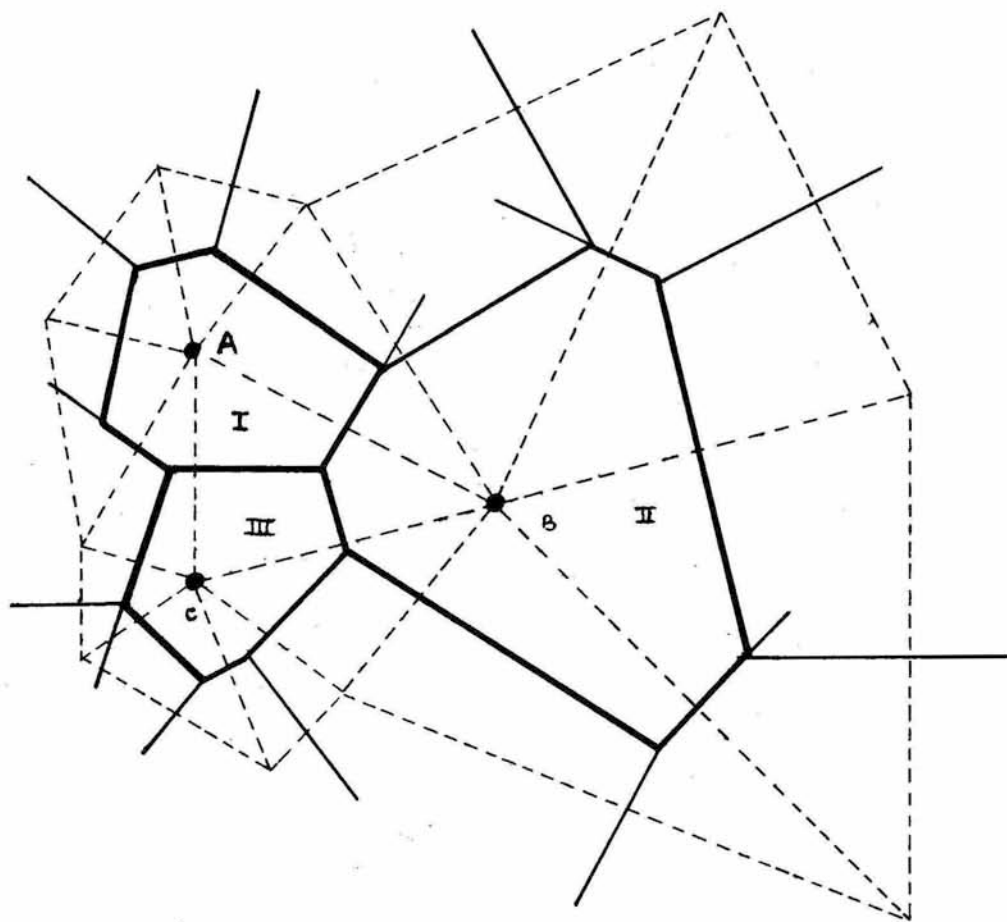
En procédant ainsi pour toutes les rubriques retenues, on définit un système de référence cohérent, où la position de chaque point est déterminée par la distance séparant la rubrique représentée de l'ensemble des autres rubriques. Un tel système doit permettre de fixer :

- la position des points correspondant à l'ensemble des rubriques;

- la position des points correspondant à l'ensemble des emplois existants.

En vertu, du principe d'assimilation, l'effectif de la rubrique A sera égal à la somme des éléments appartenant à la portion de plan I, l'effectif de la rubrique B à la somme des éléments appartenant à la portion de plan II, etc...

GRAPHIQUE I : Détermination graphique des régions du plan  
correspondant aux observations





D'une façon générale, classer les emplois selon une nomenclature donnée, revient à répartir les effectifs salariés entre différentes portions d'espace. Le découpage dépend du nombre de rubriques retenues et de la distance qui les sépare entre elles. Il s'appuie donc sur une certaine conception de l'hétérogénéité du travail, visant à créer des classes d'amplitude variable selon l'importance relative accordée aux catégories d'emplois désignées par les rubriques.

Dans ces conditions, chaque case polygonale représente directement les régions du plan correspondant aux observations, dont le nombre pourra être matérialisé par une colonne ayant la case considérée comme base et de volume proportionnel au nombre d'observations. Si l'on admet que les cases puissent devenir de plus en plus petites, les colonnes tendront à définir par leur base supérieure une surface continue, qui constitue une représentation parfaite du phénomène étudié. Le schéma théorique pur s'appuie dès lors sur une graduation continue des distances, permettant de concevoir une correspondance possible entre chaque point et un emploi réel. La discontinuité des observations, liée essentiellement à la taille des unités traitées, tend à rendre en grande partie inopérante l'utilisation directe d'une telle procédure. Il conviendrait de procéder à un "lissage" des valeurs empiriques disponibles afin de pouvoir dégager de l'observé, la forme générale de la structure des emplois.

La rubrique est un simple repère auquel on associe un nombre désignant les emplois situés dans son voisinage immédiat. Or, un emploi donné, défini par ses caractéristiques propres, est susceptible d'être rattaché à diverses rubriques avec des probabilités variables. Il semble satisfaisant d'admettre que la probabilité de rattachement est d'autant plus forte que le contenu de l'emploi est proche du contenu de la rubrique : c'est-à-dire qu'elle dépend du degré de substituabilité de l'emploi observé à l'emploi repère. Ainsi, puisqu'une rubrique est susceptible de regrouper, avec des probabilités

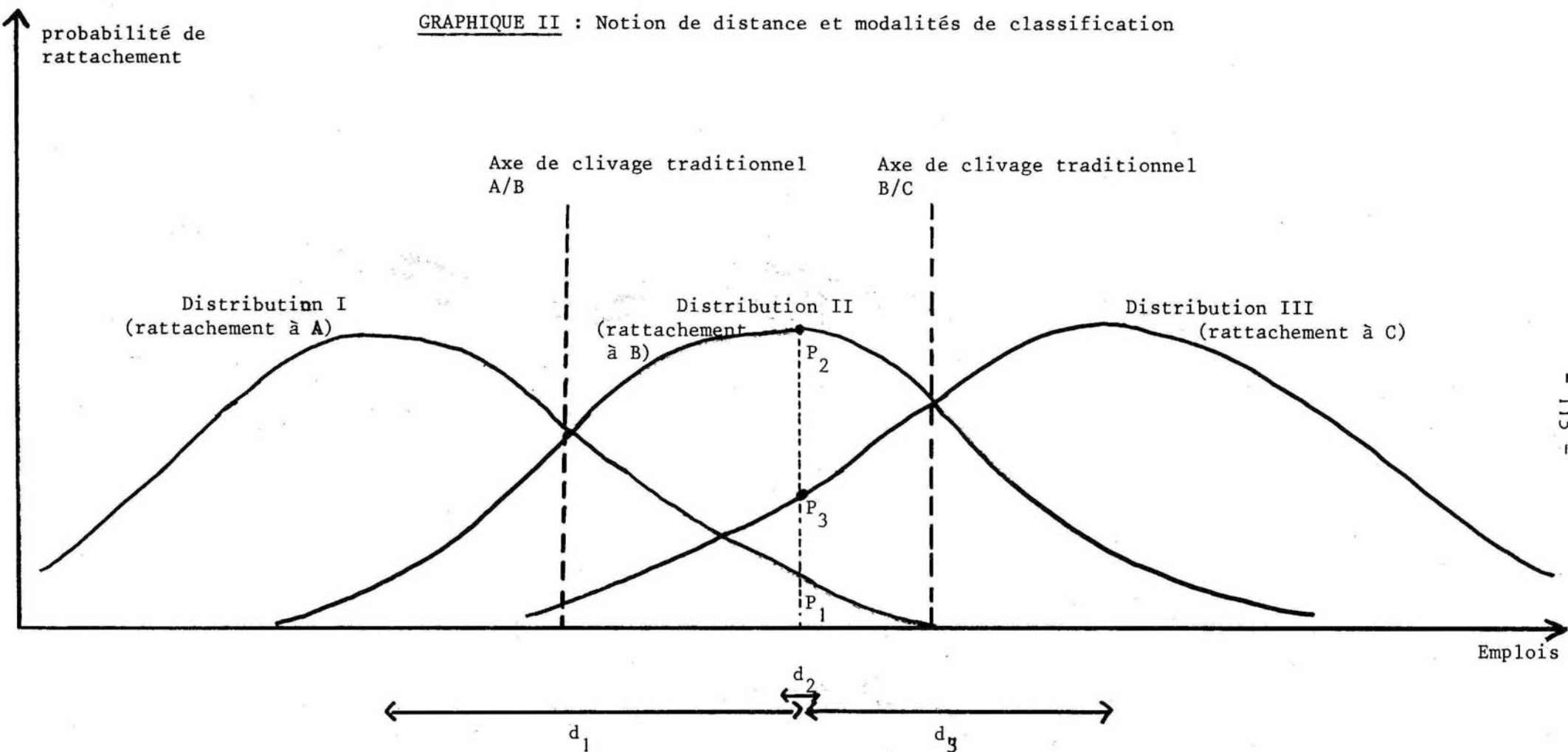
variables, diverses catégories d'emploi, en sens inverse, l'effectif répertorié sur une rubrique pourra se rapporter, avec des probabilités variables à toute une gamme d'emplois se répartissant autour du point repère.

Le concept de classification ne doit donc plus être compris dans son sens de clivage étanche : il apparaît comme un enchevêtrement de distributions de probabilité (voir graphique II), dont la combinaison, liée au choix des rubriques retenues et aux possibilités de substitution exprimées en fonction de la distance, définit les modalités de repérage des emplois. Soient  $p_1 = f(d_1)$ ,  $p_2 = f(d_2)$ ,  $p_3 = f(d_3...)$  la probabilité de rattachement d'un emploi X aux rubriques A, B, C... ; l'expression :

$$p_1 \text{ EFF (A) } + p_2 \text{ EFF (B) } + p_3 \text{ EFF (C) } + \dots$$

mesurera la densité de probabilité associée à l'emploi X. D'une façon générale, si chacune des distributions de probabilité est affectée de la proportion d'effectif observée pour la rubrique, leur combinaison linéaire permettra de construire la surface continue représentative de la structure des emplois. Dans ces conditions, comparer deux structures revient à comparer deux surfaces, c'est-à-dire à apprécier l'importance des volumes différentiels délimités par les deux surfaces structurelles.

GRAPHIQUE II : Notion de distance et modalités de classification



*B - Les justifications économiques de la démarche*

Dans sa formulation la plus simple, la démarche proposée s'apparente aux procédés d'ajustement mécanique par la méthode des moyennes mobiles. D'une façon générale, le problème posé est le suivant : on dispose d'une série d'observations présentant un caractère discontinu et irrégulier, et l'on se propose d'ajuster les valeurs disponibles afin d'éliminer tout ce qui peut raisonnablement être considéré comme secondaire par rapport au phénomène étudié.

La discontinuité des observations tient ici à la nature même du système de représentation traditionnel qui conduit, au prix d'une opération d'abstraction, à assimiler des emplois diversifiés à une rubrique ponctuelle qui les regroupe. L'irrégularité est liée d'abord aux erreurs d'observations, provenant pour l'essentiel des modalités d'affectation à une rubrique qui s'appuient dans certains cas sur un choix préférenciel ambigu. Elle s'exprime en outre et surtout, par rapport à un schéma conceptuel idéalisé, selon lequel la répartition des emplois serait le reflet d'un nécessaire équilibre organisationnel appelé à harmoniser de proche en proche les effectifs répertoriés.

A une rubrique donnée est associé un nombre qui correspond, pour chaque unité d'observation, établissement ou entreprise, à l'effectif salarié rattaché aux emplois décrits. Les comparaisons portant sur ce terme doivent être interprétées en tenant compte des glissements éventuels de classification et des écarts enregistrés sur les autres termes de comparaison. Avec le système de représentation adopté traditionnellement, les distorsions structurelles mises en évidence, qui font ressortir toutes les différences brutes indépendamment de leur signification relative, risquent de privilégier le rôle des facteurs explicatifs secondaires et donc de masquer l'influence réelle des facteurs explicatifs fondamentaux.

Les procédures d'ajustement visent principalement à éviter un tel écueil en substituant à la série des valeurs observées une série de valeurs théoriques s'inscrivant dans un cadre tendanciel général. Elles trouvent ici une justification supplémentaire liée au concept d'unité organisationnelle. La signification économique de la procédure de substitution adoptée, sous contrainte d'équilibre global, apparaît en effet de façon assez nette dans le cas présent.

Le "lissage" des valeurs observées en fonction des autres données s'appuie sur une hypothèse de compensation possible des écarts, positifs et négatifs, entre les points représentatifs et la surface théorique qui les ajuste. Cela revient seulement à admettre que, dans une unité de production, un emploi donné pourrait éventuellement être "remplacé" par un emploi voisin sans pour autant modifier sensiblement le mode organisationnel initial, l'importance de la transformation structurelle étant liée à la distance séparant les deux emplois substitués.

Ainsi se précise l'idée selon laquelle, dans le rapprochement des deux structures, relatives à deux établissements ou à un même établissement au cours du temps, l'interprétation d'une différence brute entre les effectifs d'une rubrique n'est pas indépendante de l'importance et du sens des différences entre les effectifs des autres rubriques qualitativement proches. Dans cet esprit, admettre qu'un écart positif puisse être partiellement compensé par un écart négatif, c'est seulement reconnaître que les deux catégories d'emplois désignées sont en réalité partiellement substituables et que la source de distorsion éventuelle présente un caractère mineur. L'ajustement mécanique ne fait alors que confirmer l'existence d'un phénomène économique réel, consacrant la nature du lien qualitatif entre les emplois.

Dans ces conditions, les comparaisons portant sur deux structures "ajustées" seront uniquement révélatrices des distorsions structurelles profondes susceptibles d'exister, la démarche utilisée procédant automatiquement à une sélection des valeurs significatives.

### 1) L'échelle quantitative des distances

La démarche adoptée conduit à définir la notion de classe selon une modalité d'appartenance probabiliste, le nombre théorique rattaché à la rubrique représentant alors la fréquence théorique de la classe ainsi définie. De façon concrète, l'emploi d'une telle procédure repose sur l'utilisation de deux types de relations complémentaires, dont l'établissement ne va pas sans une marge d'arbitraire importante. Il convient en effet d'abord de construire une échelle des distances permettant de hiérarchiser les rubriques entre elles. De cette échelle des distances seront ensuite déduits, par l'intermédiaire d'une formule de transformation, les coefficients de pondération adoptés.

En ce qui concerne le premier point, si la notion de distance correspond à une idée communément admise, il faut bien reconnaître que les critères appelés à l'expliciter ne s'imposent généralement pas à l'utilisateur et à l'analyste avec la netteté convaincante souhaitable. On considèrera que deux emplois sont peu distants lorsque les individus qui les occupent sont "facilement" transférables de l'un à l'autre. Dans le cas contraire, la distance sera grande, l'échelle adoptée devant principalement refléter l'importance des qualifications complémentaires requises pour permettre le passage éventuel d'un emploi à l'autre. Les critères les plus souvent proposés s'appuient dès lors sur la notion de vitesse de passage, qui peut être réellement observée en se référant aux pratiques des entreprises, ou mieux, être établie théoriquement en fonction des modalités optimales d'acquisition de la formation complémentaire nécessaire.

Dans cet esprit, une quantification de l'échelle des distances, qui préciserait la relation initiale purement qualitative, apparaît possible, les appréciations concernant la différence de

contenu des emplois décrits étant susceptibles d'admettre une mesure chiffrée. A l'heure actuelle cependant, il faut déplorer l'absence totale de résultats concrets en ce domaine ; le problème est seulement posé, mais sa résolution, d'un intérêt capital, fait encore partie des projets de recherche futurs. A défaut de meilleures données, nous avons été amenés à construire une échelle grossière, dont la justification réside essentiellement dans une analyse des principaux critères ayant servi à l'élaboration de la nomenclature. En jouant sur les moyens utilisés par le producteur d'information, il semblait en effet possible d'approcher les résultats qui auraient pu être fournis par application d'une autre méthode.

Ainsi, bien que l'objectif poursuivi soit le même, la démarche adoptée diffère dans son principe de celles qui sont le plus souvent proposées : à des critères reposant sur une certaine conception de la finalité de l'étude privilégiant le lien formation-emploi, nous avons substitué des critères plus simples à repérer, car transparaissant à travers la liste des emplois fournie, et plus simples à manier, car reflétant l'esprit ayant présidé à l'élaboration de la nomenclature. Une telle échelle des distances doit cependant, dans la mesure où ses modalités de construction dépendent en grande partie des convictions de son auteur, être considérée comme une simple hypothèse de travail, sur laquelle il est toujours possible et souhaitable de revenir.

Son but est de rendre compte du double mécanisme suivant :

- un emploi donné, appartenant à une rubrique, est susceptible de se substituer à n'importe quel autre emploi dans un taux s'exprimant en fonction inverse de la distance séparant les emplois désignés ;

- n'importe quel emploi, appartenant à une rubrique quelconque est susceptible de se substituer à l'emploi repéré dans un taux s'exprimant en fonction inverse de la distance séparant les emplois désignés.

Il s'agit là d'une généralisation de la procédure de construction des classes traditionnelles. L'opération d'abstraction visant à assimiler des emplois diversifiés à un point repère, centre de classe, exprime en effet l'idée d'une substituabilité supposée parfaite à l'intérieur de la rubrique s'opposant à une absence de substituabilité entre les rubriques. Pour rendre compte de la continuité du phénomène, une classe doit pouvoir couvrir potentiellement tout l'espace des emplois, la probabilité de lui appartenir dépendant, pour un emploi donné, de son degré de substituabilité à l'emploi représentatif du centre de la classe.

## 2) La notion de proximité

D'une façon générale, toutes les rubriques existantes peuvent être classées en fonction de leur distance à une rubrique donnée. A chaque distance peut être associé un nombre exprimant la proximité entre les emplois désignés. La relation fonctionnelle liant distance et proximité n'est cependant pas immédiate : elle dépend des critères retenus pour apprécier les deux concepts. Si l'on se réfère dans le premier cas à la possibilité de transfert et dans le deuxième cas à la possibilité de remplacement, l'un étant lié aux modalités réelles ou potentielles du passage, l'autre étant lié aux conditions autorisant la substitution éventuelle, on conçoit la nécessité d'une analyse destinée à mettre en évidence le rapport entre formation complémentaire requise et contenu spécifique des emplois.

Un emploi peut se définir comme un ensemble de tâches assumées par son titulaire ou susceptibles de l'être. Ces diverses tâches mettant en oeuvre un ensemble d'aptitudes caractéristiques, il est clair, si l'on admet un lien étroit entre formation et aptitude, que les conditions de substituabilité dépendront des possibilités de passage. Mais cela ne fait que confirmer une évidence selon laquelle distance et proximité s'expriment en relation inverse, sans pour autant fournir les éléments nécessaires à la détermination de l'échelle des



correspondances. En l'absence de support justificatif concret, le choix arbitraire d'une formule de transformation risque selon les cas d'estomper les distorsions structurelles réelles par un jeu de compensation abusif, ou de masquer les distorsions structurelles profondes si les taux de substitution adoptés sont trop faibles.

Le schéma qu'il conviendrait de construire est du type suivant. Considérons une rubrique et disposons sur un axe les points représentatifs des autres rubriques en fonction de leur distance à la rubrique repère adoptée comme origine. La relation fonctionnelle entre distance et proximité permet de faire correspondre à chaque point représentatif un taux de substituabilité tel que, si la distance à la rubrique origine est grande, le taux doit être faible et inversement. Le principe de détermination de ce taux peut découler soit de l'observation des pratiques réelles de substitution dans les entreprises, soit de l'analyse du chevauchement éventuel des distributions de tâches potentielles formant le contenu théorique de chaque emploi existant.

La notion de proximité repose en effet sur la constatation d'une certaine plasticité dans les modalités de division du travail. Il peut donc s'agir, selon le premier type d'approche, de montrer comment, sous contraintes techniques associées à un objectif donné, des structures d'emploi nettement différenciées peuvent effectivement coexister dans les entreprises, consacrant ainsi l'existence d'un phénomène de substitution possible entre emplois, dont l'expression probabiliste serait spontanément issues de l'observation. Plus généralement, une analyse du contenu même des emplois doit permettre, en remontant aux aptitudes requises, d'associer à chaque emploi une distribution de tâches potentielles susceptibles d'être accomplies par son titulaire. Dans cet esprit, à des emplois parfaitement différenciés peut correspondre une gamme d'aptitudes similaire, définissant des profils d'emploi théoriques extrêmement voisins. Selon ce deuxième

type d'approche, le taux de substitution serait issu d'une étude comparative des distributions théoriques obtenues.

Les degrés de spécialisation des modes organisationnels observés constituent une excellente illustration du phénomène. La réponse à un flux de commandes implique en effet une série d'interventions diversifiées dont le découpage en emplois, combiné au processus technologique, détermine un type d'organisation socio-technique possible. Selon les cas, la contribution de chaque individu peut être volontairement limitée à un petit nombre d'interventions ou au contraire couvrir un réseau d'interventions relativement large. Cela ne signifie pas nécessairement l'existence d'une opposition fondamentale entre les modes organisationnels choisis, la structure réelle des emplois étant susceptible, par un jeu de substitution, de masquer une structure de tâches commune ou semblable.

### 3) La distance entre établissements

La démarche adoptée doit en définitive, après analyse du système de repérage et de collecte de l'information destinée à mieux comprendre sa signification réelle et à mieux dominer ses possibilités d'utilisation concrète, permettre la construction d'un outil statistique directement opératoire ayant la double propriété d'insérer les données brutes de départ et de combiner les critères économiques sous-jacents. Au-delà du simple problème de réaffectation des emplois à leur place d'origine (celle qu'ils auraient dû occuper en l'absence de regroupement), il s'agit bien davantage de proposer un système de représentation structurelle plus réaliste que le système traditionnel car reflétant une certaine souplesse au niveau de la division du travail, une possibilité d'échange, de substitution entre emplois dont la prise en compte matérialise l'expression fondamentale du mode organisationnel choisi.

En outre, le système proposé évite les imperfections liées au mode de représentation traditionnel. La rigidité du principe d'exclusion, selon lequel un emploi ne peut être rattaché qu'à une seule rubrique indépendamment de sa distance aux autres est fortement atténuée dans le cas présent par l'introduction du concept de proximité, défini comme une possibilité de substitution entre emplois. L'hétérogénéité des classes traditionnelles, dont l'amplitude relative, liée au principe d'exclusion, dépendait du choix des rubriques retenues et de leur distance entre elles, disparaît également au profit de classes ayant une assise probabiliste parfaitement homogène. Enfin, le problème de la représentativité de l'échantillon descripteur, sur lequel planait une incertitude impossible à lever, est lui-même considérablement estompé, le mécanisme de redistribution des observations sur l'espace des emplois réalisant selon une modalité probabiliste l'opération inverse de celle du regroupement.

Pour un établissement ou une entreprise donnée, la transformation des données structurelles brutes de départ par combinaison linéaire des distributions de probabilité des rubriques, chacune des distributions étant affectée d'un poids correspondant à la fréquence observée pour la rubrique, fournit un mode de représentation en continu où la côte associée à un emploi exprime son degré d'insertion au contexte organisationnel. Plus précisément, un emploi réel regroupe des tâches diversifiées, combine des interventions multiples, réalisant un type d'adaptation concret, premier élément d'un découpage possible des activités. Les modalités de substitution font rejeter la conception d'un découpage permanent, rigide et conduisent à admettre l'existence de systèmes équivalents, mais reposant sur une combinaison différente. Dans ces conditions, le mode organisationnel réel devient un simple référenciel, image projetée d'un ensemble plus vaste dont on cherche à appréhender le contenu : c'est l'objectif poursuivi par la transformation du système de représentation.

Il est dès lors clair que les comparaisons portant sur deux surfaces continues théoriques tendront à privilégier les distorsions structurelles profondes. Certes, si la répartition des salariés selon la grille d'emplois initiale est la même, les deux volumes délimités coïncideront parfaitement et la distance entre les deux structures sera nulle. Mais, pour des structures traditionnelles différentes, la distance, mesurée par l'importance quantitative des volumes différentiels, dépendra des modalités qualitatives d'expression de cette différence. Elle pourra selon les cas, être fortement atténuée par un jeu de compensation utilisant les mécanismes de substitution entre emplois, ou au contraire ne pas être estompée si les causes de divergence sont très significatives. L'échelle des distances entre établissements doit ainsi refléter l'existence d'un phénomène économique ramené à sa juste dimension.

En se référant à un schéma bâti sur le même modèle que le système de représentation des emplois, la position de chaque établissement étant définie par la position des autres établissements et la distance qui les sépare entre eux, il est possible d'obtenir une configuration cohérente des données structurelles comparatives, point de départ d'une méthode générale d'analyse des modes organisationnels et de leur transformation.

## II - LES COMPARAISONS INTER-STRUCTURELLES

Les modalités de la division du travail à l'intérieur d'une entreprise ou d'un établissement résultent de la combinaison de l'ensemble des individus intervenant sur le processus technique, chacun réalisant une fraction de l'ensemble des interventions associées au fonctionnement de l'unité productive. Ainsi, la démarche explicative se situe au confluent de deux systèmes de causalité emboîtés privilégiant successivement la liaison organisation technique-organisation sociale et la liaison individus-tâches-emplois.

Il est clair en effet, que le processus technologique adopté joue un rôle déterminant dans la structuration du facteur travail. Le raisonnement le plus simpliste conduirait à admettre que l'organisation sociale et le reflet de l'organisation technique, la décomposition des opérations techniques devant justifier logiquement l'existence des postes de travail correspondants, qui induisent à leur tour la présence de nouveaux postes sans lesquels l'entreprise ne saurait fonctionner, ni préparer son avenir. Dans cet esprit, les caractéristiques du processus définissent une série de tâches interdépendantes, les caractéristiques du marché définissent un ensemble de qualifications rattachées aux individus et par suite les modalités de l'ajustement entre individus et tâches doivent déterminer la structure des emplois qui apparaît dès lors comme un état d'équilibre sous contrainte.

L'observation des faits tend sinon à contredire la pertinence d'un tel schéma d'analyse, du moins à en atténuer la rigueur au profit d'une plus grande souplesse d'adaptation. Des entreprises utilisant des processus technologiques semblables peuvent très bien présenter des structures d'emploi assez nettement différenciées, alors que les emplois peuvent être voisins avec des processus productifs différents. Avancer dans ce dernier cas que les caractéristiques des populations disponibles sont dominantes, ne constitue qu'une explication partielle dont on ne saurait se contenter si les structures associées à des processus de même nature ne sont pas en règle

homogènes. Force est de reconnaître que le chef d'entreprise dispose d'une marge de manoeuvre suffisamment importante puisqu'elle brise apparemment la cohérence du schéma d'analyse élémentaire.

En définitive, deux idées fondamentales semblent s'opposer : celle d'interdépendance des emplois dont la combinaison définit le mode organisationnel, celle de liberté offerte aux entrepreneurs pour structurer ces mêmes emplois. L'une exprime que l'ordonnancement adopté s'effectue selon une certaine logique et appelle la recherche des éléments permettant de la préciser. L'autre reflète le souci d'échapper à un cadre d'analyse trop rigide en admettant qu'une entreprise peut réellement choisir entre diverses formes de division du travail. La première est consacrée par le concept de structure-type, la deuxième par celui de dispersion. Si l'on suppose que la réunion de certaines caractéristiques au sein d'une même entreprise définit un ensemble de contraintes délimitant les possibilités réelles de structuration des emplois, cela revient à dire que le système organisationnel choisi dépend de quelques variables dont le nombre et la nature doivent être déterminés. Mais la liaison fonctionnelle ne vise pas l'association d'une structure unique à la présence des facteurs explicatifs. Le jeu des contraintes est plus souple, il permet d'envisager une multitude d'adaptations significatives différentes entre lesquelles le choix du chef d'entreprise est appelé à s'exprimer.

La complexité des observations, leur enchevêtrement difficilement extricable, résultent, semble-t-il, de la compatibilité de ces deux idées contradictoires en apparence, intervenant de façon simultanée dans les faits. Les principes méthodologiques de notre démarche en découlent. Ainsi se précise la signification du système de représentation structurelle adopté, où les deux éléments se retrouvent intimement liés. Ainsi se précise également l'esprit dans lequel les comparaisons devront être entreprises. La tâche est délicate. Si elle passe par une recherche des groupes dont les

modalités seront précisées, l'interprétation des résultats ne va pas sans une analyse critique visant à situer, outre le rôle éventuel des caractéristiques économiques, les conditions selon lesquelles les mécanismes de structuration des emplois sont appelés à se combiner. L'explication du réel est à ce prix, sa complexité réalise à la fois un obstacle et un guide méthodologique.

A - Distance inter-structurelle

Selon la démarche adoptée, la structure observée ne constituerait qu'un simple référentiel, matérialisant le résultat d'une combinaison possible des emplois associée à un processus de production défini. Dans la mesure où des emplois sont partiellement substituables entre eux, une généralisation des relations liant les éléments descripteurs retenus doit conduire à rapprocher des structures apparentes différenciées si elles recouvrent des entités au contenu voisin. Lorsqu'il en est ainsi, la transformation du système de représentation initial selon les modalités précédemment décrites, compose des ensembles théoriques dont la marge différentielle est relativement peu importante. En revanche, si les distorsions structurelles sont profondes, la transformation opérée les met parfaitement en évidence : Le mécanisme de compensation jouant dans le premier cas, fait alors place à un mécanisme inverse, celui de la complémentarité des écarts observés.

L'appréciation des volumes compris entre les deux surfaces théoriques associées aux structures de deux établissements, constitue dès lors un instrument de description des divergences structurelles qui les sépare.

La construction de l'outil fait appel aux notions de distance et de proximité entre emplois qu'elle combine en un système de représentation cohérent par rapport à la réalité décrite. La notion de distance intervient tout d'abord pour répartir les rubriques sur l'espace des emplois, afin de définir un schéma de référence où apparaisse clairement la dimension qualitative. Si les points représentatifs de deux rubriques se trouvent relativement écartés l'un de l'autre, cela signifie que la distance entre les emplois désignés est en règle générale importante. Dans le cas contraire, lorsque le passage d'un emploi d'une rubrique à un emploi d'une autre rubrique



est généralement facile, les points représentatifs sont relativement peu écartés.

La notion de proximité intervient au stade d'élaboration suivant, dont l'objectif vise à dégager de l'échantillon descripteur un ensemble harmonisé de données, où l'expression quantitative du système de repérage traditionnel est redistribuée sur tout l'espace des emplois. Le principe de la démarche semble s'imposer tout naturellement si l'on observe que la classification opérée requière un regroupement d'emplois, celui-ci impliquant l'idée que les éléments rattachés à la rubrique sont en fait distribués autour d'un point repère, centre de classe. Les modalités de réalisation de la procédure dépassent toutefois largement le simple cadre de la réaffectation des emplois à leur place d'origine : en mettant en lumière les conditions de substitution entre emplois, sur lesquelles repose la souplesse caractérisant la division du travail, elles doivent permettre de générer la forme probabiliste de la structure des emplois, expression fondamentale de sa forme réelle.

Si par rapport à ce schéma d'analyse, la distance entre deux structures apparaît, dans sa version la plus évidente, comme un écart brut entre volumes représentatifs, plusieurs remarques peuvent cependant être formulées :

- la mesure est synthétique : elle rend compte des distorsions structurelles dans leur ensemble, indépendamment de leur modalité d'expression ;

- la mesure est quantitative : elle fournit une information sur l'intensité des différences structurelles existantes, indépendamment de l'importance des transformations susceptibles de les résorber ;

- la mesure est statique : elle décrit un écart entre des situations organisationnelles, indépendamment de l'influence de

la structure observée sur les modalités de son évolution future.

Ces diverses caractéristiques, délimitant la portée immédiate de l'outil élaboré, appellent des développements complémentaires, destinés à préciser ses possibilités d'utilisation dans le cadre d'une recherche plus complète.

#### 1) Distance inter-structurelle et modalités de représentation

La mesure de la distance inter-structurelle telle qu'elle a été définie précédemment, exprime de façon globale l'intensité des distorsions structurelles existantes. Si la distance séparant deux établissements A et B est plus importante que la distance séparant deux établissements A et C, cela signifie que les structures d'emplois relatives aux établissements A et B diffèrent plus profondément que les structures d'emplois relatives aux établissements A et C. L'information est certes essentielle, mais elle ne précise en aucune façon la direction, le sens, les modalités d'expression de cette différence.

Pour ne pas avoir à procéder à l'analyse complexe des catégories d'emplois expliquant la composition de l'écart, nous nous référerons à un système de représentation des établissements inspiré du schéma déjà utilisé pour construire la répartition des rubriques sur l'espace des emplois. La démarche est donc du type suivant :

- la position d'un établissement quelconque A est fixée arbitrairement;
- par rapport au point représentatif de l'établissement A, la position d'un établissement quelconque B correspondra à un point de la boule de centre A et de rayon égal à la distance inter-structurelle (A, B) ;
- par rapport aux points A et B, la position d'un

établissement C correspondra à un point de l'intersection des boules centre A - rayon (A, C) et centre B - rayon (B,C) ;

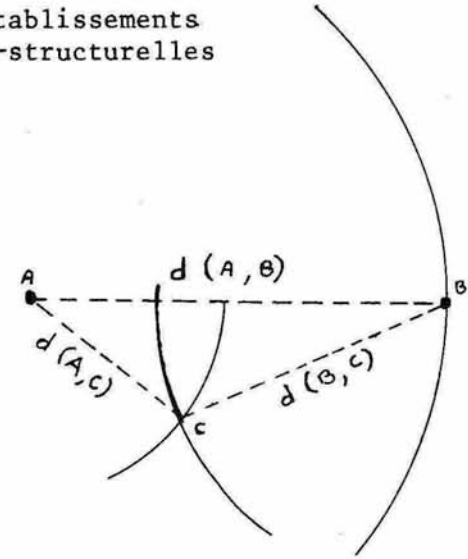
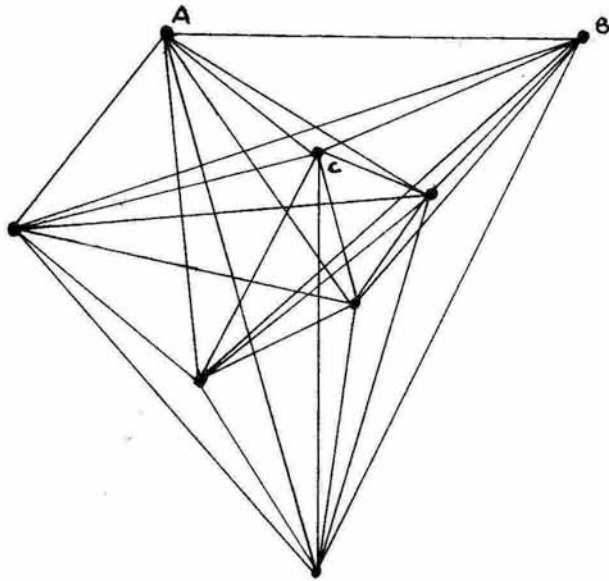
- et ainsi de suite, la position de chaque établissement étant déterminée par la position et la distance des autres établissements.

La combinaison des données issues de l'analyse renseignée dès lors, par l'intermédiaire du système de représentation, non seulement sur l'importance des distorsions structurelles observées (longueur des arêtes), mais encore sur les différences d'expression de ces distorsions (direction des arêtes) : ainsi, pour des distances voisines à un point de référence donné, l'ouverture de l'angle formé par les arêtes sera selon les cas, faible si les modalités de divergence sont semblables, ou au contraire large si elles diffèrent fortement. Un tel système permet en sens inverse de localiser les zones où la concentration des établissements est grande et donc de procéder éventuellement à une partition des structures si l'analyse consacre l'existence de groupes significatifs. Le but de l'opération consisterait à mettre en évidence des sous-ensembles structurels homogènes par rapport au critère de distance retenu, afin de dégager du nuage formé par tous les points représentatifs des établissements, un nombre réduit de noyaux structurels.

L'analyse de la population est alors considérablement simplifiée. Elle s'oriente autour de deux pôles principaux : l'identification des groupes d'abord, précisée par leur nombre, leur importance quantitative et la distance qui sépare leurs centres ; l'étude du contenu des groupes ensuite, effectuée en fonction de critères statistiques pour caractériser le degré d'homogénéité des structures composantes, et de critères économiques pour apprécier la portée explicative de certains indicateurs sur le mode organisationnel.

Afin de rendre la démarche plus exhaustive, il est souhaitable de générer, à partir de l'échantillon des observations, la

GRAPHIQUE III : Système de représentation des établissements  
en fonction des distances inter-structurelles



distribution théorique des probabilités d'apparition structurelle qui la coiffe. Une telle généralisation ne constitue pas seulement un affinement statistique de la méthode. Elle trouve sa justification dans les fondements même de l'analyse du contenu des emplois : les structures observées apparaissent en effet comme une suite de cas particuliers où les modes organisationnels s'inscrivent à l'intérieur d'une gamme théorique de choix possibles plus ou moins probables. Si l'on suppose, et l'hypothèse peut facilement être acceptée comme approximation satisfaisante, que la probabilité de réalisation décroît avec la distance à la structure réellement adoptée, la répartition initiale peut significativement être remplacée par une fonction densité dans tout l'espace des structures. C'est cette fonction qui caractérise au mieux la forme du nuage de points. C'est elle qui se prête au mieux à l'analyse des résultats observés.

## 2) L'homogénéité des critères d'analyse

Les développements qui précèdent amènent à considérer, pour employer une image destinée à faciliter la compréhension, que tout point réel, correspondant à une structure observée, répartit en quelque sorte sa propre masse dans son voisinage selon une modalité décroissant avec la distance. Ainsi, à un point quelconque de l'espace de base sera associée une densité de probabilité calculée en fonction inverse de la distance qui le sépare de l'ensemble des points représentatifs des structures réelles. Il est clair dans ces conditions que la densité de probabilité sera forte dans les zones où la concentration des établissements est elle-même relativement importante et qu'elle sera faible dans le cas contraire. Nous nous trouvons dès lors en présence d'une surface théorique continue, reflet significatif de la distribution des structures empiriques, offrant, nouvelle image, un relief plus ou moins accidenté, dont les sommets correspondent notamment au repérage des noyaux structurels susceptibles d'exister.

En pratique, l'utilisation d'une telle démarche reste entièrement subordonnée au choix du critère de distance inter-structurelle et aux modalités de détermination de l'ensemble probabilisé rattaché à la structure observée. A défaut d'une meilleure analyse, il est certain que l'introduction d'hypothèses raisonnées fournit sans nul doute un élément d'amélioration notable des résultats recherchés. Il convient cependant de s'interroger sur la signification des mécanismes économiques mis en oeuvre.

Le passage à une représentation continue du nuage de points observé a déjà été justifié et nous n'y reviendrons pas si ce n'est pour rappeler qu'il s'appuie sur un principe de substituabilité entre structures, généralisation du principe de substituabilité entre emplois. L'idée sous-jacente reste donc la même que celle qui préside à la confection de l'outil de base servant à l'appréciation des structures. Une telle homogénéité dans la démarche semble s'imposer naturellement et l'on peut se demander si, dans le cadre de la détermination de la distance inter-structurelle, le souci de logique reste sauvegardé jusqu'au bout.

Le concept de distance inter-structurelle se réfère à un mode de représentation structurelle faisant intervenir de façon combinée les notions de distance et de proximité entre emplois. Ce support théorique trouve sa justification propre en sa capacité de décrire la réalité de façon cohérente et complète. Comparer deux structures revient donc à comparer les deux surfaces continues qui leur sont associées. Le problème qui se pose à ce niveau est de savoir justement comment comparer les deux surfaces structurelles.

En termes statistiques, la réponse est claire. L'écart quantitatif, matérialisé par l'importance des volumes différentiels, admet une mesure dont la signification est liée à la validité du système de description des emplois. Si cet écart est faible, les structures comparées désignent des modes organisationnels au contenu

voisin : en effet, lorsqu'il en est ainsi, les différences observées terme à terme entre l'effectif des rubriques sont elles-mêmes relativement faibles (appréciation quantitative) et leurs modalités d'expression sur l'espace des emplois tendent à se compenser (appréciation qualitative). C'est la raison pour laquelle on peut considérer que l'instrument est parfaitement adapté à la recherche de noyaux structurels.

Mais, si la technique est appelée à fournir de bons résultats dans le cadre du rapprochement de structures entre elles, l'indice désigne-t-il réellement une distance entre structures ? La distance qui sépare des emplois est généralement définie en se référant à la notion de vitesse de passage : deux emplois sont d'autant plus proches qu'est court le temps de passage, réel ou potentiel, de l'un à l'autre. De la même façon, lorsque deux structures A et B diffèrent, ne conviendrait-il pas de s'intéresser à la transformation éventuelle de la structure A en la structure B, ou de la structure B en la structure A ? Si cette transformation est "facile", les deux structures peuvent être considérées comme étant très voisines. Si cette transformation requière tout un bouleversement du système productif, les deux structures peuvent alors être considérées comme étant très distantes l'une de l'autre.

### 3) La distance entre emplois et la distance entre structures

Quel sens donner à cette facilité ou cette complexité de transformation structurelle ? Trois orientations sont possibles, privilégiant successivement la donnée organisationnelle, la donnée emploi et la donnée évolution.

Selon le premier type d'orientation, la structure des emplois ne doit être comprise que comme élément partiel du mode organisationnel réel. Celui-ci réalise un système socio-technique complet, où, processus technologique et forme de division du travail se combinent, s'équilibrent, s'harmonisent en un jeu de mécanismes

d'inter-relations plus ou moins précises, apparaissant comme autant de contraintes techniques ou économiques, sociales ou psychologiques qui délimitent la marge de choix laissée à l'entrepreneur et qui confèrent au changement sa véritable dimension. Dans cet esprit, la distance inter-structurelle, expression de la complexité du passage d'une structure à l'autre, non seulement fait intervenir le temps de passage entre emplois, mais encore doit englober, sous forme de pondérations, la chaîne des causalités réciproques.

Est-ce là le véritable objectif poursuivi ? S'il s'agit de quantifier la distance entre deux structures d'emplois, les seules variables à prendre en considération sont des variables d'emplois. Le passage d'une structure à une autre, terme de comparaison retenu, requière, si les structures sont différentes, un minimum de transferts d'emplois, dont la réalisation apparaît d'autant plus facile que la distance entre emplois transférés est faible. Le système de représentation adopté pour décrire la structure se prête alors parfaitement à une telle détermination, puisqu'il reproduit en même temps que la répartition des emplois, l'échelle des distances qui les séparent entre eux.

Dans sa formulation la plus simple, la démarche consistait à définir essentiellement l'importance relative des effectifs à transférer. Les volumes différenciels couvrent en effet les excédents et les déficits structurels locaux qui, par construction, se compensent nécessairement, les premiers désignant une masse à répartir, les deuxièmes appelant la répartition de cette masse pour réaliser l'adéquation éventuelle. En revanche, les modalités du transfert étaient négligées et peu importait notamment que les zones excédentaires soient séparées des zones déficitaires par une grande distance ou au contraire par une distance réduite.

A l'heure actuelle, on considère que le déplacement d'une masse  $m$  sur une distance  $d$  nécessite, pour employer une image destinée à faciliter la compréhension, une énergie  $W(d, m)$ , fonction de la distance à parcourir et de la masse à transférer. Les structures



d'emplois sont représentées par des mesures de probabilité sur l'espace des emplois muni d'une métrique : soit  $P$  la distribution relative à l'établissement  $A$ ,  $P'$  la distribution relative à l'établissement  $B$ ,  $P$  et  $P'$  étant de même masse totale. On se demande alors qu'elle est "l'énergie" minimale à dépenser pour passer de la distribution  $P$  à la distribution  $P'$ . Ce problème classique admet une solution, qui est fourni par le minimum de la fonction selon laquelle se distribue la probabilité de transition de  $P$  à  $P'$ .

La démarche proposée apparaît dès lors comme une généralisation, à travers un cadre descripteur cohérent, de l'ensemble des principes analytiques mis en oeuvre. Son gros inconvénient est qu'elle nécessite l'utilisation d'une procédure de calcul, certes simple dans sa conception, mais dont la répétition systématique pour tous les couples d'établissements risque de se heurter à une lourdeur quelque peu excessive. En outre, si la détermination de la distance inter-structurelle fait bien intervenir toute la trame des distances entre emplois, l'échelle obtenue n'est valide que par rapport aux critères utilisés, qui définissent à la fois sa portée, sa signification et les modalités de son utilisation.

Souhaitons-nous en particulier rapprocher les établissements selon leur affinité évolutive, alors l'ordonnancement élaboré, reposant sur une comparaison des structures d'emplois, peut être remis en cause, dans la mesure où l'état initial s'avèrerait peu explicatif des mécanismes de transformation structurelle réellement observés.

*B - Un cadre de recherche explicative*

Le système de représentation de l'ensemble des structures observées, reposant à la fois sur les concepts de distance et de substituabilité inter-structurelles, généralisation des concepts de distance et de substituabilité entre emplois, décrit de façon ordonnée et claire une réalité complexe difficilement saisissable de prime abord : il constitue ainsi un outil d'analyse destiné à faciliter les recherches explicatives. Un des résultats escomptés serait de pouvoir associer un "type structurel" à une gamme de variables caractérisant la vie de l'entreprise ou de l'établissement, ou, plus généralement d'établir une relation probabiliste entre les formes de division du travail et la valeur de quelques indicateurs économiques fondamentaux.

L'idée maîtresse est plus séduisante que facile à mettre en oeuvre, selon laquelle l'ensemble des possibilités offertes aux entreprises s'inscrit en pratique à l'intérieur de limites assez étroites, bornées par un jeu de contraintes relativement rigides qui régiraient tant les modalités d'organisation que les modalités d'évolution tendancielle. En d'autres termes, le chef d'entreprise n'est pas libre de faire ce qu'il veut, comme il le veut, il définit sa stratégie en se référant à une situation existante qui précise son cadre d'action et au milieu environnant qui infléchit sa politique de croissance. L'importance du subi modèle en quelque sorte la plage de liberté qui lui est accordée, compose la forme de la loi de comportement dont relèvent ses possibilités d'évolution, et, en définitive, le système adopté et les mécanismes qui président à son changement, prennent une signification particulière où le déterminisme se dégage peu à peu de la loi des grands nombres.

Dans cet esprit, la structure observée pour une unité de production devrait pouvoir être rattachée à un type structurel dont l'agencement correspond à une combinaison déterminée de valeurs caractéristiques. Ou, en inversant le problème, la présence, dans un secteur d'activité, d'entreprises ayant des structures d'emploi homogènes, devrait permettre d'expliquer l'influence des contraintes économiques sur le mode organisationnel. Le premier objectif est donc

de faire ressortir de la masse des données les groupes qui la composent. Il faut repérer les zones de forte densité, celles qui réunissent dans leur voisinage une proportion relativement élevée d'entreprises, présentant des structures d'emploi assez peu différenciées. L'instrument d'analyse mis en oeuvre répond, semble-t-il, convenablement à de telles préoccupations et fournit par là même une intéressante direction de recherche explicative.

Son efficacité dépend toutefois du degré de signification du clivage obtenu. La méthode repose en effet, dans ses modalités d'application, sur un principe de substitution justifié théoriquement, mais dont la mesure est, dans l'état actuel des connaissances nécessairement arbitraire. L'échelle construite peut, selon les cas, écraser les ruptures d'homogénéité ou multiplier le nombre de groupes, générer une fonction densité unimodale ou faire apparaître autant de zones que d'observations. Faire une partition, c'est choisir entre ces deux extrêmes, puisque la souplesse de l'outil le permet. Encore faut-il que la solution retenue distingue des zones suffisamment distantes les unes des autres et regroupe des structures suffisamment homogènes entre elles. La première condition est nécessaire si l'on veut éviter un problème de chevauchement d'influences difficilement isolables. La deuxième condition est nécessaire si l'on veut définir une association caractéristiques-structure pertinente. Avec des centres de zone séparés par une faible distance, il est naturel de s'attendre à une extrême confusion des caractéristiques dégagées, reflétant la rencontre de structures voisines dans des entreprises de nature différente. Avec des observations se répartissant sous la forme d'un pseudo-continuum sans rupture franche, il devient impossible de se prononcer directement sur les variables recherchées : leurs influences jouent alors vraisemblablement de façon combinée, la structure observée participant ainsi d'un mélange plus ou moins complexe de lois explicatives.

Ces difficultés ne sont-elles pas inhérentes à la nature même du phénomène étudié ? Ne découlent-elles pas logiquement des principes d'analyse mis en oeuvre ? Il est important de le préciser pour ne pas bloquer irrémédiablement devant un mur d'incompréhension où la simplicité souhaitée fait place à une réalité différente. L'obstacle est-il lui-même franchissable ? La théorie permet-elle de poser le problème en des termes acceptant une solution statistique satisfaisante ? Des réponses à ces questions dépend la progression de nos travaux.

### 1 - Analyse économique et observation statistique

La référence analytique à des structures-type, qui apparaîtraient comme autant de modèles théoriques associés à un ensemble de contraintes économiques, ne constitue que le point de départ d'un raisonnement plus complexe visant à décrire une réalité extrêmement diversifiée dans ses formes et dans ses mécanismes évolutifs.

Le problème doit être posé en des termes permettant l'adaptation des principes méthodologiques mis en oeuvre au mode de représentation des observations réelles. D'un côté, nous disposons d'un schéma descriptif, dans lequel la distribution des structures est matérialisée par une surface continue. La construction d'un tel schéma s'appuie sur une réflexion relative au contenu des notions de structure et de distance inter-structurelle, conduisant à transformer l'ensemble des informations brutes en un système de représentation cohérent où sont étroitement combinées données quantitatives et appréciations qualitatives, données ponctuelles et expressions probabilistes. D'un autre côté, nous avons recours à un schéma explicatif, selon lequel la distribution des structures pourrait être interprétée par l'existence de structures-type, dont le nombre et le contenu apparaissent comme les inconnues du problème à résoudre.

On sait seulement que la construction d'un tel schéma doit permettre de réaliser la meilleure adéquation possible au schéma descriptif précédemment établi.

Plusieurs démarches peuvent être envisagées. La première consisterait à tester la validité de certaines hypothèses fondées sur des analyses partielles et étant susceptibles de devenir opérantes dans un cadre généralisé. On a pu voir notamment que le mode organisationnel n'était pas indépendant du type de produit fabriqué, de la nature du client auquel il était destiné, et du processus de fabrication utilisé, qu'il existait une liaison non négligeable entre les modalités de la division du travail et la taille, combinées à d'autres caractéristiques d'entreprise, comme le choix des techniques mises en oeuvre. Il devient dès lors possible d'établir une classification a priori des entreprises, s'appuyant sur des critères d'autant plus pertinents que leur assise théorique est confirmée par l'analyse. Dans cet esprit, le concept de structure-type apparaît comme la structure moyenne des entreprises composant le groupe. Son potentiel explicatif dépend individuellement du degré d'homogénéité observé et globalement du degré d'adéquation à la réalité décrite.

Plus précisément, la création de groupes d'entreprises, résultat d'une partition opérée en fonction de la possession de certaines caractéristiques, a pour vocation principale la mise en évidence de modes organisationnels semblables lorsque certaines conditions sont réunies, différents lorsqu'elles ne le sont pas. L'idée se réfère donc davantage à la notion d'organisation, configurée par l'organigramme de l'entreprise, qu'à celle de structure, plus restrictive, désignant une distribution quantitative des emplois. Deux conséquences en résultent : autour du pôle fixé par le modèle organisationnel se distribue la marge de liberté laissée aux entrepreneurs pour structurer les emplois, expliquant l'existence d'un phénomène d'attraction compatible avec l'idée de dispersion; la possession de caractéristiques hétérogènes (entreprise fabriquant par exemple plusieurs types de produits, s'adressant simultanément à une clientèle industrie et grand public) autorise la référence multiple à des modèles organisationnels différents, chacun corres-

pondant à des morceaux de l'organigramme complet, expliquant l'existence d'un phénomène structurel complexe compatible avec l'idée de structure-type.

Ainsi, en supposant connues les caractéristiques définissant le mode organisationnel, l'échantillon des observations, après sélection des entreprises présentant une homogénéité parfaite par rapport aux caractéristiques retenues, fournit un cadre de détermination des structures-type et de la dispersion liée à la marque de manœuvre disponible. En définitive, à un jeu de critères significatifs correspond une distribution de structures plus ou moins probables, à une combinaison de jeux de critères significatifs correspond une combinaison similaire des distributions de probabilité composantes. Il est dès lors possible de construire un schéma explicatif d'ensemble, où la traduction des caractéristiques d'entreprises en terme de structure se matérialise globalement sous la forme d'une surface continue directement comparable à celle issue du schéma purement descriptif.

## 2 - Choix d'une unité de raisonnement

La démarche directe qui vient d'être présentée suppose au préalable connus les critères de clivage et donc que le problème puisse être immédiatement résolu malgré la pauvreté des connaissances acquises en la matière. Or, ni le choix des caractéristiques explicatives, ni le choix de l'unité de base appelée à servir de support au raisonnement ne semblent s'imposer nettement. L'approche est d'autant plus délicate que le système auquel on se réfère, entreprise ou établissement, peut constituer un ensemble composite, offrant à l'analyse un enchevêtrement de modalités d'explications partielles qui se combinent, se chevauchent, risquent de s'estomper en se combinant ou de s'amplifier en s'ajoutant.

On sait seulement que dans bien des cas, l'établissement

représente une entité tronquée par rapport à l'entreprise qui réalise alors une fusion de fonctions interdépendantes, un regroupement de services complémentaires étant en pratique séparés géographiquement. La logique de l'analyse commande donc de comparer des structures d'entreprises dans la mesure où l'établissement est susceptible de dissocier des séries de tâches étroitement liées entre elles. Mais, si la définition générale de l'entreprise repose sur un concept d'unité selon lequel un ensemble de moyens sont mis en oeuvre pour satisfaire à un objectif économique déterminé, la mécanique combinatoire n'est pas simple : verticalement d'abord, où les enchaînements fonctionnels autorisent un choix diversifié dans les modalités de division du travail; horizontalement ensuite, où la pluralité des caractéristiques de fonctionnement autorise la juxtaposition éventuelle de systèmes organisationnels plus ou moins indépendants les uns des autres.

La structure des emplois dans une entreprise peut dès lors apparaître comme une addition de structures partielles qui correspondraient chacune à une des distributions d'emplois susceptibles d'être rattachées à un mode organisationnel spécifique, la référence simultanée à des modes organisationnels différents étant justifiée quant à elle par la réunion possible de caractéristiques multiples au sein de l'entreprise. La notion de structure partielle, qui constitue la clé de voûte du schéma d'analyse proposé, ne doit pas être confondue avec une **simple** décomposition en services ou en fonctions, l'interdépendance des éléments composants étant alors évidente bien que généralement imprécise. Le principe de la méthode est inverse, puisqu'il s'agit de repérer le plus petit ensemble structurel susceptible d'avoir une existence autonome. Dans nombre d'entreprises, cet ensemble structurel élémentaire correspondra directement à la structure totale qui réalise un équilibre fonctionnel indissociable. Mais chaque fois que les liaisons entre catégories d'emplois s'ordonnent simultanément autour de plusieurs axes différents, force est de reconnaître que l'analyse de l'équilibre global apparent procède d'un double mécanisme d'interdépendance et de **juxtaposition**. La structure observée pour l'entreprise résulte dans ces conditions d'une combinaison de structures **élémentaires** qu'il est nécessaire d'identifier, la construction d'un schéma explicatif d'ensemble

faisant appel à autant de schémas explicatifs partiels qu'il existe de distributions composantes.

L'analyse de la structure des emplois dans une entreprise conduit donc en quelque sorte à remplir les cases d'un tableau synthétique, présenté ici à titre d'illustration, où mécanismes et données s'inscrivent en accord avec les principes méthodologiques mis en oeuvre. Les colonnes acceptant une ventilation du personnel sont celles qui correspondent aux systèmes de référence adoptés. Dans le cas le plus simple, la structure observée apparaît directement comme une réalisation de la distribution de probabilité des structures associées à un seul mode organisationnel-type. Plus généralement, elle apparaît comme une somme de réalisation des distributions de probabilités associées aux divers modes organisationnels intervenant.

Tels sont les principes sur lesquels semble devoir reposer l'analyse. En permettant de mieux appréhender les modalités d'expression de la liberté offerte aux entrepreneurs pour structurer les emplois, ils fournissent une direction de recherche destinée à lever l'apparente contradiction entre une réalité observée qui est complexe et des éléments d'explication qui pourraient être fondamentalement simples. L'existence de quelques types organisationnels, en nombre supposé réduit a priori, est en effet parfaitement compatible avec l'extrême dispersion des situations observées; elle en est même la cause logique pour peu que les mécanismes présentés soient significatifs.

### 3 - Procédure d'analyse

La démarche conceptuellement la plus simple consisterait dès lors à rechercher simultanément structures-type et structures partielles, le seul critère d'appréciation étant celui de la meilleure adéquation possible du schéma d'analyse au schéma descripteur. Les hypothèses sous-jacentes sont les suivantes : on suppose que dans le cadre d'un secteur d'activité, les systèmes d'organisation concevables sont en nombre limité, qu'une entreprise peut se référer conjointement à



CADRE D'ANALYSE DE LA STRUCTURE DES EMPLOIS DANS UNE ENTREPRISE

Fonction	Rubrique	MODES ORGANISATIONNELS - TYPE					m	Structure observée
		1	2	3	4			
1	1	$x_{11}$	$x_{12}$	$x_{13}$	$x_{14}$	.....	$x_{1m}$	$X_{1.}$
	2	$x_{21}$	$x_{22}$	$x_{23}$	$x_{24}$	.....	$x_{2m}$	$X_{2.}$
	3	$x_{31}$	$x_{32}$	$x_{33}$	$x_{34}$	.....	$x_{3m}$	$X_{3.}$
2	4	$x_{41}$	$x_{42}$	$x_{43}$	$x_{44}$	.....	$x_{4m}$	$X_{4.}$
	5	$x_{51}$	$x_{52}$	$x_{53}$	$x_{54}$	.....	$x_{5m}$	$X_{5.}$
	6	$x_{61}$	$x_{62}$	$x_{63}$	$x_{64}$	.....	$x_{6m}$	$X_{6.}$
	7	$x_{71}$	$x_{72}$	$x_{73}$	$x_{74}$	.....	$x_{7m}$	$X_{7.}$
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
	n	$x_{n1}$	$x_{n2}$	$x_{n3}$	$x_{n4}$	.....	$x_{nm}$	$X_{n.}$
Structure en composantes partielles		$X_{.1}$	$X_{.2}$	$X_{.3}$	$X_{.4}$	.....	$X_{.m}$	$X_{..}$

plusieurs systèmes différents, et que les structures partielles se distribuent autour de leur moyenne selon une certaine loi de probabilité. Quels sont les grands types de structuration du facteur travail, comment se décompose la structure réelle de chaque entreprise en structures partielles, quelle fonction distribue les écarts entre structures partielles et modèles de référence ? La réponse à l'une de ces questions dépend statistiquement de la réponse apportée aux deux autres, ce qui implique la définition d'un algorithme de calcul complexe où les mécanismes d'interaction deviennent prépondérants. La solution d'ensemble doit en outre avoir un sens, être significative, ce qui implique l'adjonction de critères de vraisemblance économique à des critères purement statistiques. L'opération peut paraître hasardeuse à plusieurs titres. En ce qui nous concerne, nous retiendrons qu'elle préjuge à l'avance de la nature du résultat à obtenir. Le raisonnement, en effet, s'appuie sur un jeu d'hypothèses, fait appel à des modalités explicatives, qui constituent autant d'éléments à mettre en évidence, avant de confectionner le moule assemblant réalité observée et mécanismes sous la forme d'un schéma synthétique.

Une étape préalable est nécessaire, qui vise à approfondir les connaissances disponibles, à préciser des idées, certes séduisantes, puisqu'elles concilient simplicité interprétative et complexité des données, mais dont le fondement relève plus de l'extrapolation de concepts théoriques que de l'observation des faits. Les approches traditionnelles n'étant jamais parvenues à leur fin, leur orientation pouvait selon les cas apparaître irréaliste ou insuffisamment développée. La supposer irréaliste, c'était s'engager dans une voie de recherche excluant toute influence notable des caractéristiques économiques et techniques sur le mode organisationnel, pour faire place à un système où la liberté de choix, exprimée sans contrainte apparente, autorise la plus grande diversité des situations concrètes. Il faut bien reconnaître que les exemples abondent dans ce sens et que l'observation des faits tient lieu

d'argument jusqu'à preuve du contraire, malgré l'insatisfaction théorique éprouvée par l'analyste. Mais cette discordance entre des principes qui semblent s'imposer et une réalité qui semble les rejeter, ne constitue-t-elle pas le meilleur stimulant d'une recherche plus approfondie, dont l'objectif est de montrer comment les uns parviennent à expliquer l'autre? Il peut s'agir, à défaut d'avoir trouvé les "bonnes" variables, de poursuivre l'exploration systématique des causalités potentielles. Il peut s'agir, à défaut d'avoir trouvé un cadre d'explication opérant, de poursuivre les tentatives de décomposition de la structure totale d'entreprise en structures partielles.

Face à cette voie de recherche, deux orientations sont possibles. La première consiste à affiner, par l'intermédiaire d'enquêtes spécifiques, le domaine des connaissances acquises, dans l'espoir de répondre simultanément aux interrogations relatives aux variables explicatives et à l'unité de raisonnement optimale. C'est le même but que nous poursuivons, mais à partir d'une méthode plus directement axée sur l'observation statistique. N'est-il pas possible de dégager de l'enchevêtrement des données un embryon de résultat qui, réinjecté dans le processus, servirait à alimenter le système de compréhension final? Au stade initial, il s'agit donc seulement de repérer les groupes structurels, formés par le rapprochement de structures totales, et d'analyser les caractéristiques qui s'y rattachent. Mais ce point de départ s'accompagne d'une prise de conscience de l'insuffisance de la démarche traditionnelle. Il doit être prolongé, l'espoir étant ici de se procurer à chaque étape une partie des chaînons devant composer l'ensemble explicatif complet. Ainsi se précise le cadre d'analyse proposé, qui appelle la construction progressive d'un système complexe à l'échelle de la complexité du réel. Reposant sur un schéma théorique, il vise à montrer sa validité à travers l'observation des faits, reposant sur l'observation, il vise à montrer sa logique à travers un ensemble de lois explicatives. La justification de notre démarche réside dans la nécessité d'une telle confrontation.

*C H A P I T R E   I V*

---

*MODELE D'EXPLORATION DES DISTORSIONS INTERSTRUCTURELLES*

*FORMALISATION MATHEMATIQUE DU PROBLEME*

Nous allons reprendre dans ce chapitre les définitions des emplois et des structures, telles qu'elles ont été décrites dans les chapitres précédents afin d'en donner une formalisation mathématique qui nous permettra de procéder à une analyse de l'ensemble des structures. Nous essayerons de plus de mettre en évidence les paramètres que nous introduisons et de donner un moyen de les interpréter.

Pour ne pas faire deux versions de ce chapitre, l'une destinée à un public peu informé des notions mathématiques, l'autre plus technique dans laquelle nous aurions été amenés à répéter de larges extraits de la première, nous avons mis un trait en marge en face des passages plus techniques. Ces passages peuvent ne pas être lus sans que cela nuise à la compréhension générale du chapitre. Ils sont néanmoins nécessaires à ceux qui voudront suivre les détails techniques des hypothèses et du raisonnement.

Pour décrire la structure des emplois dans une unité industrielle (entreprise ou établissement), nous disposons du nombre de salariés de l'unité, dans chaque poste d'une nomenclature donnée. Le problème que nous proposons de traiter est le problème général de l'analyse des données : comment classer ces structures d'emploi, comment les regrouper de la façon la plus pertinente possible en un nombre restreint de sous-ensembles, comment caractériser ces sous-ensembles et les situer les uns par rapport aux autres ?

Dans la première partie de cet exposé, nous étudierons une méthode de comparaison des structures, c'est-à-dire un calcul d'une distance entre vecteurs-structures, dans la seconde partie, nous nous intéresserons au classement des structures.

I - COMPARAISON DES STRUCTURES D'EMPLOI

A) Exposé théorique des problèmes

1) Aspects généraux

a) Choix d'un système de référence

Notre problème est ici d'établir une distance entre structures d'emploi, celles-ci étant représentées dans une nomenclature donnée par le nombre (ou le pourcentage) de salariés qui figurent dans chacun des postes. Les formules les plus généralement utilisées tiennent compte soit de l'écart type  $\sigma$ , soit de la moyenne  $f$  des valeurs trouvées dans la population dans chaque poste. On rapporte en effet les valeurs absolues des différences à l'une ou l'autre de ces quantités en écrivant, si  $X = (x_i / i = 1, n)$  et  $Y = (y_i / i = 1, n)$  sont deux structures exprimées en  $n$  rubriques :

$$d_{\sigma}^2(X,Y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{\sigma_i^2}$$

ou

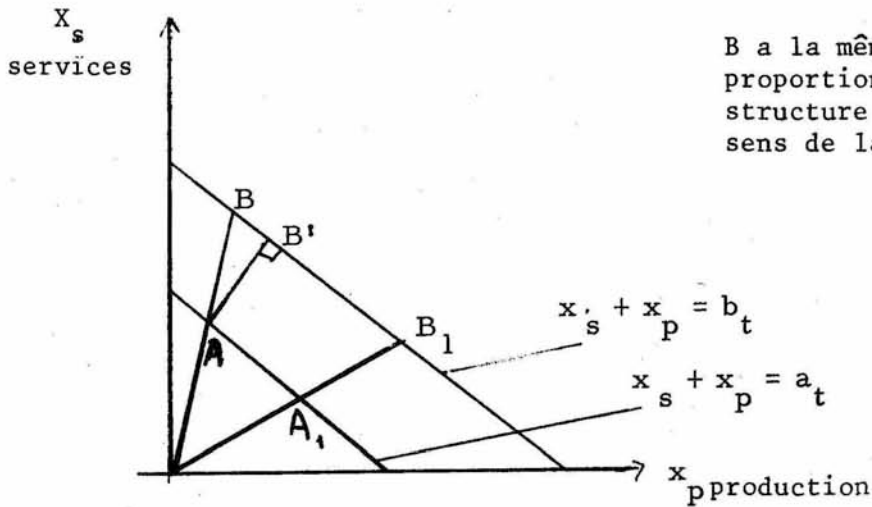
$$d_f^2(X,Y) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{f_i}$$

On reconnaît dans la seconde expression, la formule utilisée en analyse factorielle des correspondances.

Notons tout de suite que l'utilisation de telles formules sur des vecteurs structures exprimés en nombre de salariés réels, et non en proportions, présente dans le cas qui nous intéresse, au moins une incohérence, que nous démontrons l'exemple suivant : supposons une nomenclature en deux postes : service et production. Soit une unité de structure A composée de  $a_{\Delta}$  employés dans les services et  $a_{\rho}$  dans la production avec un total de  $a_{\epsilon} = a_{\Delta} + a_{\rho}$  salariés. Considérons alors l'ensemble des unités employant au total  $b_{\epsilon}$  salariés,  $b_{\epsilon}$  étant différent de  $a_{\epsilon}$ , supposons le supérieur par exemple. Si nous cherchons dans l'ensemble B des structures correspondant à ces unités, celle qui est la plus proche de A au sens d'une distance du type considéré, on ne trouvera pas dans le cas général

la structure ayant les mêmes proportions que A dans les services et dans la production, c'est-à-dire, qu'on ne trouvera pas la structure  $B = (b_s, b_p)$  telle que  $\frac{b_s}{b_p} = \frac{a_s}{a_p}$

On voit l'illustration de ce résultat sur le graphique ci-dessous :



Un autre inconvénient est l'effet d'écartement des grandes unités dû à l'homothétie. Supposons par exemple 4 structures A, A<sub>1</sub>, B, B<sub>1</sub> telles que les couples (A,B) et (A<sub>1</sub>, B<sub>1</sub>) aient les mêmes proportions de salariés dans les deux postes et les couples (A, A<sub>1</sub>) et (B, B<sub>1</sub>) le même total de salariés. Du fait de la différence de taille, la distance entre B et B<sub>1</sub> est beaucoup plus grande que celle entre A et A<sub>1</sub>, ce qui n'est pas forcément justifié.

Les remarques que nous venons de faire restent valables pour un nombre de postes supérieurs à deux et pour toute distance mesurée par une forme quadratique quelconque. Comme nous resterons dans ce cadre, il nous a semblé plus logique d'employer une distance qui soit décomposée en deux termes dont le premier mesure un écart entre les structures exprimées en proportions et le second, une différence de taille.

b) Expression de la distance

Dans la pratique, nous n'avons pas fait intervenir la différence des tailles. Nous dirons néanmoins quelques mots de l'utilisation d'un tel terme. Comme pour toutes les distances, il faut que ce terme soit de la forme  $f(|a_t - b_t|)$  où  $f$  est une fonction croissante et convexe. La fonction la plus utilisée est  $d^2 = \beta (a_t - b_t)^2$ . Le problème qui se posera alors sera de déterminer les importances relatives à donner à ce terme et à celui qui mesure l'écart des structures. Dorénavant, les structures seront exprimées uniquement en proportions.

Soient deux structures  $A = (a_i)_{i \in I}$  et  $B = (b_i)_{i \in I}$  où  $I$  est l'ensemble d'indices de la nomenclature utilisée et une distance de la forme :

$$d^2(A, B) = \sum_{i \in I} \alpha_i (a_i - b_i)^2$$

où  $\alpha_i$  est un coefficient positif. Ce type de distance a l'inconvénient d'être fortement dépendant de la nomenclature retenue. En effet, supposons que deux postes  $j$  et  $k$  recouvrent des emplois peu différents et que dans un cas on décide de les séparer et dans un second de les confondre, on aurait alors :

$$d_1^2(A, B) = \sum_{i \in I - \{j, k\}} \alpha_i (a_i - b_i)^2 + \alpha_j (a_j - b_j)^2 + \alpha_k (a_k - b_k)^2$$

$$d_2^2(A, B) = \sum_{i \in I - \{j, k\}} \alpha_i (a_i - b_i)^2 + \alpha_{j+k} [(a_j + a_k) - (b_j + b_k)]^2$$

On voit que dans le calcul de  $d_1$ , la répartition entre  $j$  et  $k$  interviendra assez fortement, alors qu'elle n'intervient pas dans le calcul de  $d_2$ . Or, si les postes  $j$  et  $k$  sont très proches, il semble plus normal que la répartition entre les postes joue peu. Les distances du type ci-dessus ne tiennent pas compte de la ressemblance entre postes de la nomenclature. Supposons par exemple qu'à partir d'une structure  $A$  on construise d'une part une structure  $B_1$  par le transfert d'une quantité  $\Delta_x$  d'un poste  $i$  (par ex. ouvrier électricien qualifié) à un poste  $j_1$  (par exemple : technicien électricien), d'autre part une structure  $B_2$  par le transfert



de la même quantité  $\Delta_x$  du même poste  $i$  à un poste  $j_2$  (par ex. : personnel administratif qualifié) beaucoup plus différent de  $i$  que  $j_1$ . Il semblera évident à un économiste (et même à un non-initié) que la structure  $B_1$  ressemble plus à la structure  $A$  que la structure  $B_2$ . Or, si les coefficients attachés aux postes  $j_1$  et  $j_2$  sont du même ordre de grandeur, ces deux structures seront aussi éloignées l'une que l'autre de  $A$ , dans une distance du type proposé, puisque l'on aura :

$$\begin{aligned}d^2(A, B_1) &= \alpha_i (\Delta_x)^2 + \alpha_{j_1} (\Delta_x)^2 \\d^2(A, B_2) &= \alpha_i (\Delta_x)^2 + \alpha_{j_2} (\Delta_x)^2\end{aligned}$$

Nous avons essayé de pallier à ces inconvénients en nous servant de formules de calcul qui semblaient mieux rendre compte de la réalité des distances interstructurelles. Pour ce faire, nous avons établi une formalisation des emplois et des structures d'emploi

## 2 - Les emplois

L'ensemble des emplois possibles ou existants a été partitionné en un nombre fini de classes, rubriques de la nomenclature utilisée pour l'enquête. Il est évident que l'utilisation d'un nombre restreint d'appellations constitue une simplification de la réalité, mais elle est nécessaire. Une rubrique de la nomenclature recouvre une distribution d'emplois réels variés.

Pour formaliser cette notion, nous sommes amenés à munir l'espace des emplois d'une structure d'espace probalisable, donc de définir une tribu de parties sur cet espace sur laquelle seront définies les probabilités liées aux rubriques de la nomenclature.

De plus des emplois réels quasiment identiques ne sont pas forcément attribués à une même rubrique en raison des erreurs de ceux qui remplissent les déclarations. La nomenclature

qui devrait idéalement être une partition de l'ensemble des emplois (recouvrement de cet ensemble par des parties deux à deux disjointes), comporte parfois des imprécisions qui laissent un certain choix à l'enquêté. Il peut arriver aussi que ce dernier interprète mal des définitions, d'où le chevauchement de certaines rubriques.

Par ailleurs, nous avons essayé de quantifier la différence entre deux emplois. Nous en verrons l'utilité pour une étude correcte des comparaisons des structures. Nous avons utilisé les critères qui nous ont semblé les plus pertinents quant aux buts de l'étude, pour établir une matrice des distances entre les rubriques de la nomenclature. Il est évident que le mode de calcul adopté n'a rien d'absolu. Nous donnerons plus loin un moyen d'interpréter l'effet des paramètres introduits (1). On peut considérer que deux rubriques se chevauchent d'autant plus que la distance calculée entre elles, qui est en fait une distance entre les moyennes, est faible.

### 3 - Les structures et les distances interstructurelles

#### a) Les structures

Les rubriques étant des distributions sur l'espace des emplois, et les structures des distributions sur l'ensemble des rubriques de la nomenclature, les structures d'emploi peuvent être considérées comme des distributions sur l'espace des emplois, en associant à chaque rubrique, non pas la distribution réelle qu'elle a dans une structure donnée, mais une distribution type, celle par exemple qu'elle aurait dans la population totale des salariés.

La distribution sur l'espace des emplois, associée à la structure, est la combinaison barycentrique des distributions associées aux rubriques avec le système de poids donné par le vecteur structure. La comparaison des structures se ramène donc

(1) Les matrices construites comportent en fait des 1 sur la diagonale, pour les transformer en matrices de distance, il faudrait appliquer aux coefficients une fonction transformant le 1 en 0, qui soit croissante et qui conserve l'inégalité triangulaire.

à l'établissement d'une distance sur les distributions de probabilité dans un espace métrique puisque nous établissons une distance entre emplois.

On a exposé au chapitre précédent une méthode de calcul de distance interstructurelle fondée sur la représentation suivante : l'espace des emplois étant assimilé à un plan, chaque rubrique représentant un continuum d'emplois admet une fonction densité. Donc une structure admet aussi une fonction densité, puisqu'elle est la combinaison barycentrique des distributions associées aux différentes rubriques. En prenant comme plan de base le plan des emplois et un axe vertical pour la valeur de la fonction densité, on peut représenter les structures par des surfaces dans cet espace à trois dimensions ou par les volumes que ces surfaces délimitent avec le plan des emplois.

On aura reconnu les mesures habituelles sur  $L^1(\mu)$  où  $\mu$  est la mesure dont on a muni l'espace des emplois.

b) la méthode de calcul des distances la plus logique

La manière la plus logique pour poser le problème des distances interstructurelles serait de le formuler comme un problème de coût, de la façon suivante :

• étant donné deux structures  $S_1$  et  $S_2$  et des coûts  $(C_j^i)$  de passage d'un emploi à un autre, fonctions croissantes des distances séparant les couples de postes  $(i, j)$ , comment passer de  $S_1$  à  $S_2$  avec un coût minimum ?

Pour que ce coût minimum puisse être une mesure de la distance entre  $S_1$  et  $S_2$ , il faut que le coût de passage de  $S_1$  à  $S_2$  soit le même que celui de  $S_2$  à  $S_1$ , quels que soient  $S_1$  et  $S_2$ . Il suffit donc que le coût de passage d'un poste  $i$  à un poste  $j$  soit le même que celui du poste  $j$  au poste  $i$ . Dans le cas présent, ce coût pourrait être interprété comme la somme du coût de la formation complémentaire à donner à un employé occupant le poste  $i$  pour qu'il soit capable d'occuper le poste  $j$ , et pour ce même employé, du coût de la formation qui lui a été donnée pour qu'il occupe le poste  $i$  et qui ne lui est plus nécessaire s'il est affecté au poste  $j$ . On voit que cette somme des coûts des formations

complémentaires et gaspillées est alors une fonction symétrique en ce sens que le coût de passage  $i$  à  $j$  est le même que le coût de passage de  $j$  à  $i$ .

Si les structures sont des probabilités sur l'espace des emplois, la formalisation générale du problème est la suivante :

. soit,  $\Omega$  l'espace des emplois,  $\omega$  un emploi,  $\mathcal{A}$  la tribu des parties de  $\Omega$ .

. soit  $m \times f(d)$  l'énergie dépensée pour déplacer une masse  $m$  d'une distance  $d$  sur l'espace métrique des emplois. La distance entre deux structures  $S_1$  et  $S_2$  est donnée par le minimum de l'intégrale :

$$\int_{\Omega} \int_{\Omega} P_1^2(\omega_1, d\omega_1) dS_1(\omega_1) \cdot f[d(\omega_1, \omega_2)]$$

sur l'ensemble des probabilités de transition de  $(\Omega \times \mathcal{A})$  dans  $\mathbb{R}$  qui transforment la probabilité  $S_1$  en la probabilité  $S_2$ .

L'expression  $dS_1(\omega_1)$  est la masse élémentaire dans l'élément d'espace  $d\omega_1$  pour la mesure de probabilité  $S_1$ .  $P_1^2(\omega_1, d\omega_2)$

est la proportion de la masse élémentaire en  $d\omega_1$  qui est transférée dans l'élément  $d\omega_2$  par la probabilité de transition  $P_1^2$ .

L'expression  $P_1^2(\omega_1, d\omega_2) dS_1(\omega_1)$  représente donc l'élément de masse qui est transféré de  $\omega_1$  à  $\omega_2$ . L'ensemble des probabilités

de transition transformant  $S_1$  en  $S_2$  n'est pas vide, puisqu'il contient au moins la probabilité de transition qui, à tout point, de  $\Omega$  associe  $S_2$ . De plus, cette intégrale est positive, donc minorée, dès qu'on suppose  $f$  positive, aussi existe-t-il un minimum.

Cependant, il ne semble pas que l'on puisse résoudre facilement le problème dans le cas où  $\Omega$  a un nombre infini d'éléments. Si au contraire on considère qu'il n'existe qu'un nombre fini d'emplois, le problème se ramène à un problème classique de programmation linéaire.

En effet dans ce cas, les probabilités de transition sont des endomorphismes linéaires transformant  $S_1$  en  $S_2$  dans le simplexe de  $R^{(A)}$  des lois de probabilité. Si  $C = (C_{ij}^j)$  est la matrice des coûts de passage du poste  $i$  au poste  $j$ , cela revient à chercher une matrice  $M(p, p)$  (avec  $p = |A|$ ) à coefficients positifs dont chaque colonne a une somme égale à 1, telle que :

$$M \cdot S_1 = S_2$$

et qui minimise l'expression

$$\sum_{i,j} m_{ij}^j S_i C_{ij}^j$$

L'ensemble des matrices  $M$  qui satisfont les contraintes n'est pas vide puisqu'il y a en particulier la matrice dont tous les vecteurs colonnes sont égaux à  $S_2$ ; de plus cet ensemble, considéré comme sous-ensemble de  $R^{m^2}$ , est contenu dans le cube unité (d'arête 0,1 sur tous les axes), il est fermé puisque les contraintes sont des contraintes d'égalité ou d'inégalité larges, donc il existe des solutions.

De toute façon, nous n'avons pas les éléments nécessaires de définition de l'espace des emplois pour résoudre ce problème dans le cas présent. L'exposé des solutions que nous venons de faire peut alors paraître superflu et ne répondre qu'à un désir abstrait et inutile de formalisation. En fait, d'une part, il nous semble plus honnête de montrer les différents aspects du problème, d'autre part, nous rencontrerons un problème absolument analogue et pour lequel nous aurons tous les éléments de définition, à propos de la description de l'ensemble des structures d'emploi. Nous demandons donc à nos lecteurs de se souvenir de ce problème et des remarques que nous avons faites à ce sujet.

Nous pourrions traiter ce problème comme si les rubriques étaient entièrement homogènes (tous les emplois d'une rubrique étant strictement analogues), mais pour des considérations techniques de temps de calcul en ordinateur, nous ne retiendrons pas cette solution.

c) La méthode retenue

Revenons à la représentation des structures que nous avons donnée à la fin du paragraphe a. En prenant comme espace de base, l'espace des emplois et en ajoutant un axe pour mesurer la densité par rapport à une mesure sur l'espace des emplois, nous avons vu qu'une structure peut être représentée par un volume que délimitent la surface d'équation, la densité  $f(\omega)$  en chaque point de l'espace des emplois de la distribution de probabilité associée à la structure, avec l'espace des emplois. Deux structures limitant des volumes distincts, nous allons prendre comme mesure de leur distance la mesure du volume différentiel de ces deux structures, c'est-à-dire de l'ensemble des volumes qui appartiennent à l'une des structures sans appartenir à l'autre.

En fait, cette mesure représente la distance entre deux fonctions densité de probabilité dans  $L^1(\mu)$  où  $\mu$  est une mesure de référence sur l'espace probabilisable des emplois. Cela suppose donc l'adjonction à la structure d'espace probabilisable de l'espace des emplois d'une mesure par rapport à laquelle toutes les distributions de probabilité associées aux rubriques admettent une fonction densité (1). Soit  $\mu$  une de ces mesures. La mesure des distances entre deux structures est alors indépendante du choix de  $\mu$ . Pour la commodité des calculs nous choisirons de mesurer les distances dans  $L^2(\mu)$  qui est un hilbert et non dans  $L^1(\mu)$ . Nous pourrions ainsi nous ramener à un produit scalaire simple. Le choix de la mesure  $\mu$  conditionne alors les valeurs des distances. Nous nous servirons d'ailleurs plus loin de ce paramètre.

Dans le cas où les rubriques seraient totalement séparées, c'est-à-dire où il n'y aurait jamais d'ambiguïté pour l'affectation des emplois, cette méthode reviendrait à faire directement la somme des carrés (ou des valeurs absolues) des différences poste à poste, entre deux vecteurs structures (les mesures associées aux différentes rubriques seraient alors étrangères deux à deux).

---

(1) cf. Bases mathématiques du calcul des probabilités Neveu (ex.IV.1.3 p.107)

Pour l'application de cette méthode, nous considérerons donc que les rubriques se chevauchent peu ou prou quelle que soit la distance de leurs moyennes et d'autant plus que cette distance est faible. De plus, nous considérerons que les distributions sont entièrement définies par leur moyenne (ce qui ne signifie pas que les distributions de probabilité associées à toutes les rubriques ont même norme dans  $L^2(\mu)$  lorsque leurs moyennes sont distinctes). Précisons notre propos par le calcul de la distance de deux structures dont la différence ne portent que sur deux postes, soit :

$$S_1 = S_2 + \Delta_i \text{ (salariés sur le poste } i) + \Delta_j \text{ (salariés sur le poste } j)$$

Les structures étant exprimées en proportions, dans un tel cas, nous aurons :

$$\Delta_j = -\Delta_i$$

Alors, si nous posons  $f_i$  densité par rapport à la mesure  $\mu$  de la rubrique  $i$ , on obtient :

$$\begin{aligned} d^2(S_1, S_2) &= \int_{\Omega} \left( \frac{dS_1}{d\mu} - \frac{dS_2}{d\mu} \right)^2 d\mu = \int_{\Omega} (\Delta_i f_i + \Delta_j f_j)^2 d\mu \\ &= \Delta_i^2 \int_{\Omega} f_i^2 d\mu + \Delta_j^2 \int_{\Omega} f_j^2 d\mu + 2\Delta_i \Delta_j \int_{\Omega} f_i f_j d\mu \end{aligned}$$

Si nous posons :  $q_i^j = \int_{\Omega} f_i f_j d\mu$   
 nous obtenons la formule :

$$d^2(S_1, S_2) = q_i^i \Delta_i^2 + q_j^j \Delta_j^2 + 2 q_i^j \Delta_i \Delta_j$$

Dans cette formule les coefficients positifs  $q_i^i$  indiquent l'importance que l'on attache à la rubrique  $i$ ; notre hypothèse sur le chevauchement des rubriques signifie que le rapport  $q_i^i / q_i^i \cdot q_j^j$  est d'autant plus proche de 1 que les rubriques  $i$  et  $j$  sont proches l'une de l'autre.

Il est toujours positif puisque  $\mu$  est une mesure positive, et inférieur à 1 d'après l'inégalité de Schwartz.

$q_{ij}^j$  donne donc une mesure de l'ampleur du chevauchement. Si  $\Delta_j$  est l'opposé de  $\Delta_i$ , alors la formule donnant la distance devient :

$$d^2(S_1, S_2) = q_{ij}^i \Delta_i^b + q_{ij}^j \Delta_i^b - 2 q_{ij}^j \Delta_i^b$$

Si les rubriques  $i$  et  $j$  sont confondues, on aura alors d'après les hypothèses déjà faites :

$$q_{ij}^i = q_{ij}^j = q_{ij}^i$$

et donc :

$$d^2(S_1, S_2) = 0$$

ce qui est bien le résultat que nous voulons.

Si au contraire les rubriques  $i$  et  $j$  sont infiniment éloignées, on peut considérer qu'elles ne se chevauchent plus et on obtient :

$$d^2(S_1, S_2) = q_{ij}^i \Delta_i^b + q_{ij}^j \Delta_j^b$$

c'est-à-dire une formule du type des distances déjà vu.

Le calcul général se fait de la façon suivante :

$$\frac{d S_1}{d \mu} = \sum_{i \in I} \Delta_i^1 f_i \quad ; \quad \frac{d S_2}{d \mu} = \sum_{i \in I} \Delta_i^2 f_i$$

$$\begin{aligned} d^2(S_1, S_2) &= \int_{\Omega} \left( \sum_{i \in I} \Delta_i^1 f_i - \sum_{i \in I} \Delta_i^2 f_i \right)^2 d\mu \\ &= \int_{\Omega} \left[ \sum_{i \in I} (\Delta_i^1 - \Delta_i^2)^2 f_i^2 \right. \\ &\quad \left. + 2 \sum_{i \in I} \sum_{\substack{j > i \\ j \in I}} (\Delta_i^1 - \Delta_i^2)(\Delta_j^1 - \Delta_j^2) f_i f_j \right] d\mu \end{aligned}$$

En supposant  $I$  ordonné.



Les intégrales qui interviennent sont toutes finies par hypothèse. I étant un ensemble fini, on peut intervertir les signes :

$$d^2 (S_1, S_2) = \sum_{i \in I} (\delta_i^1 - \delta_i^2)^2 \int_{\mathcal{N}} f_i^2 d\mathcal{N} + 2 \sum_{\substack{i \in I \\ j > i \\ j \in I}} (\delta_i^1 - \delta_i^2) (\delta_j^1 - \delta_j^2) \int_{\mathcal{N}} f_i f_j d\mathcal{N}$$

En posant :

$$q_i^j = \int_{\mathcal{N}} f_i f_j d\mathcal{N}$$

on obtient la formule générale :

$$d^2 (S_1, S_2) = \sum_{i \in I} q_i^i (\delta_i^1 - \delta_i^2)^2 + 2 \sum_{\substack{i \in I \\ j > i \\ j \in I}} (\delta_i^1 - \delta_i^2) (\delta_j^1 - \delta_j^2) q_i^j$$

ou si l'on appelle Q la matrice carrée  $(q_i^j)_{i \in I, j \in I}$  et  $(S_1 - S_2)$  le vecteur colonne différence des structures, on a la formule matricielle :

$$d^2 (S_1, S_2) = {}^t(S_1 - S_2) Q (S_1 - S_2)$$

Cette matrice Q est construite à partir d'une matrice de distance entre les postes en posant que le rapport  $q_i^j / q_i^i \cdot q_j^j$  est d'autant plus faible que les postes i et j sont éloignés. Il faut remarquer que dans le cas présent, il est peu probable que les distances entre les postes puissent être calculées comme dans l'analyse factorielle des correspondances. En effet, deux postes qui présentent de grandes similitudes quant au travail et aux qualifications qu'ils impliquent, peuvent être très écartés dans un tel calcul des distances, puisqu'il peut arriver fréquemment qu'un chef d'entreprise ne remplisse que l'une des deux catégories. La déclaration d'une seule des deux rubriques peut provenir, soit d'une assimilation abusive dans l'esprit de celui qui remplit l'enquête, et qui donc rangera dans une seule catégorie l'ensemble des salariés qui auraient dû être rattachés à l'une ou l'autre, soit parce qu'ayant effectivement des employés dans une des deux catégories, il lui aurait été superflu d'embaucher des personnes dans la catégorie voisine. Donc, il peut arriver souvent que lorsqu'une des catégories est utilisée, l'autre ne le soit pas, ce qui entrainera des profils

tout à fait différents pour ces deux postes. Néanmoins, il est peut être possible d'essayer de construire le tableau des distances interpostes à partir du tableau des données d'une manière différent plutôt que de le faire en raisonnant seulement d'un point de vue économique. Nous n'avons pas étudié la question ici.

Voyons sur un exemple comment les coefficients de la forme quadratique de distances interstructurelles pourraient être calculés à partir de la matrice des distances entre les moyennes des rubriques.

Supposons que l'espace des emplois soit  $\mathbb{R}^n$  muni de la mesure de Lebesgue et que les rubriques soient des distributions normales d'écart-type  $\sigma$  dont la densité de probabilité soit :

$$f_i(x) = \frac{1}{(\sqrt{2\pi} \sigma)^n} e^{-\frac{\|x - A_i\|^2}{\sigma^2}}$$

où  $A_i$  est le centre de la rubrique  $i$ .

Alors le coefficient  $q_i^j$  de la forme quadratique est égal à :

$$q_i^j = \frac{1}{(\sigma\sqrt{2\pi})^n} \int_{\mathbb{R}^n} \exp\left(-\frac{\|x - A_i\|^2}{\sigma^2} - \frac{\|x - A_j\|^2}{\sigma^2}\right) dx$$

Si on se réfère à un système d'axes orthogonaux dans lequel les coordonnées sont  $x = (X^k)$ ,  $k = 1, n$ , on peut décomposer l'intégrale en produit de la façon suivante :

$$q_i^j = \frac{1}{(\sigma^2 2\pi)^n} \prod_{k=1}^n \int_{\mathbb{R}} \exp\left(-\frac{(X^k - A_i^k)^2 - (X^k - A_j^k)^2}{\sigma^2}\right) dX^k$$

On met alors le polynôme en exponentielle sous la forme canonique

$$(X^k - A_i^k)^2 + (X^k - A_j^k)^2 = 2 \left[ X^k - \frac{A_i^k + A_j^k}{2} \right]^2 + \frac{(A_i^k - A_j^k)^2}{2}$$

Appelons I le milieu de  $(A_i, A_j)$ ; la formule devient :

$$q_{ij}^j = \frac{1}{(\sigma^2 2\pi)^n} \prod_{k=1}^n \exp\left(-\frac{(A_i^k - A_j^k)^2}{2\sigma^2}\right) \int_{\mathbb{R}} \exp\left[-\frac{2}{\sigma^2} (X^k - I^k)^2\right] dX^k$$

L'intégrale se calcule par translation et donne :

$$q_{ij}^j = \frac{(\sigma \sqrt{\pi})^n}{(2\pi \sigma^2)^n} \exp\left(-\frac{\|A_i - A_j\|^2}{2\sigma^2}\right) = \frac{1}{(2\sqrt{\pi} \sigma)^n} \exp\left(-\frac{\|A_i - A_j\|^2}{2\sigma^2}\right)$$

On voit apparaître l'influence de la distance entre  $A_i$  et  $A_j$  sur la valeur du coefficient  $q_{ij}^j$ .

#### d) Synthèse

La distance entre deux structures est donc mesurée par une forme quadratique :

$$d^2(S_1, S_2) = \sum_{i \in I} q_i^i (s_i^1 - s_i^2)^2 + 2 \sum_{i \in I} \sum_{\substack{j > i \\ j \in I}} q_i^j (s_i^1 - s_i^2)(s_j^1 - s_j^2)$$

où les coefficients  $q_i^j$  sont positifs et tels que :

$$q_i^{j2} < q_i^i \cdot q_j^j$$

Posons :  $\alpha_i = \sqrt{q_i^i}$  et  $\epsilon_i^j = \frac{q_i^j}{\alpha_i \alpha_j}$

l'expression de la distance devient :

$$d^2(S_1, S_2) = \sum_{i \in I} c_i^i [\alpha_i (s_i^1 - s_i^2)]^2 + 2 \sum_{i \in I} \sum_{\substack{j > i \\ j \in I}} c_i^j \alpha_i (s_i^1 - s_i^2) \alpha_j (s_j^1 - s_j^2)$$

avec  $C_i^i = 1$  pour tout  $i$ .

La matrice  $C = (C_{ij}^j)_{i \in \Gamma, j \in \Gamma}$  comporte donc des 1 sur la diagonale et des nombres positifs ou nuls inférieure à 1 ailleurs. La différence sur chaque poste est multipliée par un nombre  $\alpha_i$ .

Appelons D la matrice diagonale de termes  $\alpha_i$ . On peut décomposer la matrice Q en un produit D.C.D. Nous pouvons donner une autre interprétation de Q sous cette forme. Nous considérons que les  $\alpha_i$  sont des coefficients proportionnels aux normes des vecteurs de 1% de différence sur les postes. Autrement dit, une différence entre deux vecteurs de 1% sur un poste engendre une distance  $\alpha_i^2 \cdot k$ , où k est un coefficient constant indépendant du poste. La matrice C représente dans le repère des axes liés à chacun des postes, les cosinus entre les différents axes. Un cosinus nul signifie que les deux axes sont orthogonaux et plus le cosinus est proche de 1 plus l'angle entre les deux axes est faible. Supposons un vecteur différence dont deux composantes i et j ne soient pas nulles, on a, comme on l'a déjà vu plus haut:

$$d^2 (S_1, S_2) = C_{i,i}^i (\alpha_i \Delta_i)^2 + C_{j,j}^j (\alpha_j \Delta_j)^2 + 2 C_{j,i}^i (\alpha_i \Delta_i) (\alpha_j \Delta_j)$$

Or,  $C_{i,i}^i$  et  $C_{j,j}^j$  sont tous deux égaux à 1. L'égalité est donc :

$$\begin{aligned} d^2 (S_1, S_2) &= (\alpha_i \Delta_i)^2 + (\alpha_j \Delta_j)^2 + 2 C_{j,i}^i (\alpha_i \Delta_i) (\alpha_j \Delta_j) \\ &= (\alpha_i \Delta_i + \alpha_j \Delta_j)^2 - 2 (1 - C_{j,i}^i) (\alpha_i \Delta_i) (\alpha_j \Delta_j) \end{aligned}$$

Nous voyons que plus le coefficient  $C_{j,i}^i$  se rapproche de 1, plus la somme  $(\alpha_i \Delta_i + \alpha_j \Delta_j)$  prend une valeur relative importante dans la valeur de la distance, vis à vis du second terme que l'on peut considérer comme un terme de répartition entre les deux postes. A la limite, lorsque le coefficient  $C_{j,i}^i$  est égal à 1, seul le terme  $(\alpha_i \Delta_i + \alpha_j \Delta_j)$  compte, c'est-à-dire que seule la valeur globale sur les deux postes réunis intervient dans le calcul. (Il faut bien sûr, que tous les coefficients relatifs aux postes i et j soient alors égaux

pour tous les postes  $k$  et que  $\alpha_i = \alpha_j$ . Les postes  $i$  et  $j$  sont alors confondus.

Nous nous servirons plus loin de cette décomposition de la matrice  $Q$ .

## B - Les implications de la méthode

### 1) Une contrainte

Puisque nous mesurons les distances entre les structures par une forme quadratique, il faut s'assurer que cette méthode définit bien une distance, c'est-à-dire que quels que soient les couples  $(S_1, S_2)$  de structures distinctes, la distance calculée est bien strictement positive. En effet, supposons qu'il existe deux structures  $S_1$  et  $S_2$  distinctes dans la nomenclature utilisée et qu'elles soient séparées par une distance nulle. Alors cela entraînerait que si deux structures sont séparées par un vecteur proportionnel au vecteur  $S_1 - S_2$ , elles sont aussi à une distance nulle l'une de l'autre. Cela signifierait que l'ensemble des salariés des postes affectés d'un signe positif et l'ensemble des salariés des postes affectés d'un signe négatif sont exactement interchangeables, dans les proportions indiquées par le vecteur  $S_1 - S_2$ . Dans une nomenclature où l'on considérerait que tous les postes sont distincts, ce fait serait difficilement justifiable, donc, il faudra s'assurer que la matrice  $Q$  est bien celle d'une forme quadratique positive. Comme cette matrice est par construction symétrique réelle, elle admet une base de vecteurs propres, il suffit donc de s'assurer que les valeurs propres sont toutes strictement positives. Nous avons tenté de définir des conditions sur les coefficients de la matrice pour que celle-ci représente une forme définie positive. Nous n'avons pas abouti.

Pour une démonstration directe, nous avons pensé à la formalisation suivante :

. si la propriété est vraie sur la matrice  $C$  définie elle est vraie pour  $Q$ . Nous raisonnerons sur la matrice  $C$  considérée comme une matrice de cosinus entre des axes dont

chacun représenterait un poste de la nomenclature. Il faut et il suffit alors que ces vecteurs ainsi définis par les angles forment une base. Dans le cas de trois axes, cela s'exprime simplement en notant que chacun des angles doit être inférieur à la somme des deux autres et supérieur à la valeur absolue de leur différence, ce qui donne la relation :

$$|\text{Arc cos } a_1^2 - \text{Arc cos } a_1^3| < \text{Arc cos } a_2^3 < \text{Arc cos } a_1^2 + \text{Arc cos } a_1^3$$

Il faudrait rendre cette formule symétrique et la généraliser. Il pourrait être intéressant de procéder par récurrence, en trouvant comment à partir d'une matrice  $C_{n-1}$ , satisfaisant la condition, on peut construire par addition d'une ligne et d'une colonne une matrice  $C$  qui satisfasse, elle aussi, la condition.

Comme théorème général pouvant être appliqué à ce problème, nous n'avons trouvé que le théorème d'Hadamard qui impose que le terme diagonal soit supérieur à la somme des termes non diagonaux (dans une matrice à coefficients positifs) dans chaque ligne ou colonne. Cette condition est trop forte pour que nous l'utilisions étant donnée la façon dont nous voulons présenter les regroupements de postes.

Toutefois, les matrices que nous avons construites jusqu'à présent satisfont toutes à la contrainte, nous respectons (inconsciemment) les conditions qui peuvent exister. Cela nous engage à penser que ces dernières doivent être simples et naturelles.

## 2 - Proposition pour une interprétation

Supposons que deux distances  $d_1$  et  $d_2$  aient été établies pour mesurer les écarts entre structures. Pour les comparer, nous proposons de rechercher le système de vecteur  $(V_1, \dots, V_i, \dots)$  tel que étant donné un vecteur différence de structures  $(S_1 - S_2)$  on puisse écrire :

$$S_1 - S_2 = \sum_{i \in I} x_i V_i$$

et :

$$d_1^2(S_1, S_2) = \sum_{i \in I} C_1^i x_i^2$$

$$d_2^2(S_1, S_2) = \sum_{i \in I} C_2^i x_i^2$$

On sait qu'une telle base orthogonale pour les deux matrices  $Q_1$  et  $Q_2$  existe toujours.

Le passage de la distance  $d_1$  à la distance  $d_2$  se traduit alors pour un vecteur  $V_i$  par un effet multiplicateur  $\frac{C_2^i}{C_1^i}$

Il est donc intéressant d'interpréter les vecteurs  $V_i$  qui subissent le plus grand effet multiplicateur et ceux qui subissent le plus petit effet multiplicateur.

L'interprétation d'un vecteur  $V$  de composantes  $(V^1 \dots V^I)$  se fera en notant et en interprétant l'opposition entre les groupes de postes ayant une forte composante négative et de ceux ayant une forte composante positive. Pour faciliter l'interprétation, on pourra regrouper les vecteurs ayant des coefficients multiplicateurs sensiblement égaux et rechercher dans le sous-espace qu'ils engendrent, le système de vecteurs le plus facilement interprétable. Celui-ci peut être par exemple le système dont les différents vecteurs ont un maximum de composantes nulles.

Il faudra évidemment chercher les bases orthogonales de ces sous-espaces, l'interprétation directe des termes rectangles étant plus délicate.

Si on appelle  $Q_1^{-1/2}$ , une racine symétrique réelle carrée de  $Q_1$ , la recherche de la base à la fois  $Q_1$ - et  $Q_2$ -orthogonale revient à la recherche des vecteurs propres de  $Q_1^{-1/2} Q_2 Q_1^{-1/2}$  comme nous allons le montrer rapidement. Notons tout d'abord que  $Q_1^{-1/2}$  existe et est réelle puisque  $Q_1$  comme matrice symétrique réelle est diagonalisable. Soit  $(X_1 \dots X_n)$  le système de vecteurs propres de  $Q_1^{-1/2} Q_2 Q_1^{-1/2}$ . Ces vecteurs sont  $Q_1^{-1/2} Q_2 Q_1^{-1/2}$  orthogonaux deux à deux et ils sont de plus I-orthogonaux (I étant la matrice unité) puisque  $Q_1^{-1/2} Q_2 Q_1^{-1/2}$

est symétrique réelle. Alors le système  $(\bar{Q}_1^{-1/2} X_1 \dots \bar{Q}_1^{-1/2} X_n)$  répond à la question. En effet  ${}^t(Q_1^{-1/2} X_i) Q_2 (Q_1^{-1/2} X_j)$  est (pour  $i \neq j$ ) nul à cause de la première orthogonalité du système  $(X_i)$  et  ${}^t(Q_1^{-1/2} X_i) Q_1 (Q_1^{-1/2} X_j) = X_i I X_j$  est nul à cause de la seconde orthogonalité.

Nous avons exposé plus haut les raisons de l'introduction d'une matrice non diagonale. Cette interprétation des effets de la matrice nous permettra de voir dans quelle mesure nous sommes arrivés à l'effet voulu.

Dans le cas où l'on veut mesurer les effets de l'introduction d'une matrice  $Q$  par rapport à une matrice diagonale  $\Lambda$ , la méthode revient à chercher les vecteurs propres de la matrice de distance qui forment une base orthogonale pour les deux systèmes et à comparer les rapports des valeurs propres à la masse des vecteurs par la matrice diagonale, qui sont les valeurs propres du produit  $\Lambda^{-1/2} Q \Lambda^{-1/2}$  où  $\Lambda^{-1/2}$  est une racine carrée de  $\Lambda$ .

Dans le cas où les structures sont exprimées en proportions, les vecteurs différence de structures sont dans l'hyperplan  $H$  d'équation  $\sum_i x_i = 0$  dans le repère ayant comme axes les postes, tous normés de la même façon.

Si le vecteur  $(1 \dots 1)$  n'est pas vecteur propre du produit  $Q_2^{-1/2} Q_1 Q_2^{-1/2}$ , il n'existe pas, dans  $H$ , de base à la fois  $Q_1$  et  $Q_2$  orthogonale.

On peut cependant former une base de  $H$  dont  $(n - 2)$  vecteurs forment un système orthogonal pour les deux systèmes de distances en prenant par exemple dans  $H$  le vecteur subissant le plus grand effet multiplicateur dans le passage de  $Q_1$  à  $Q_2$ , puis dans l'intersection de dimension  $(n-3)$  des supplémentaires par rapport à  $H Q_1^-$  et  $Q_2^-$  orthogonal à



ce vecteur, le vecteur subissant le plus grand effet multiplicateur et ainsi de suite. On pourra choisir comme dernier vecteur, soit un vecteur  $Q_1$  ou  $Q_2$  orthogonal au système ainsi formé, soit un vecteur qui minimise certains termes rectangles dans les deux systèmes.

### 3 - Importance ou poids des différents postes

Dans le calcul d'une distance entre deux structures, les écarts pour chacun des postes ne sont pas identiques et l'on peut souvent dire que la différence porte sur quelques postes dont les écarts dans les deux structures sont importants par rapport aux autres.

Si par exemple, on calcule les distances de la façon la plus simple par :

$$d^2 (S_1, S_2) = \sum_{i \in I} (s_i^1 - s_i^2)^2$$

il est probable que les postes d'ouvriers qualifiés et spécialisés qui sont beaucoup plus fournis que les autres créeront la plus grande partie de la distance. Nous voulons dire par là, que si l'on ne considérait que ces deux postes et que l'on représentait chacune des structures par un point dans un plan dont les axes perpendiculaires les représenteraient, on aurait une bonne approximation des distances réelles. Or, si l'on s'astreint à prendre en considération tous les postes, c'est parce que, en général, on désire ne pas tomber dans une telle approximation. En analyse factorielle des correspondances, on considère par exemple comme nous l'avons déjà vu, la distance :

$$d^2 (S_1, S_2) = \sum_{i \in I} \frac{(s_i^1 - s_i^2)^2}{f_i}$$

où  $f_i$  est la moyenne des  $s_i$ . Dans ce cas, on défavorise donc par rapport à la première forme, les postes les plus fournis et on favorise les postes les moins fournis.

Examinons plus précisément comment nous pouvons quantifier cette notion.

. si l'on considère une distance de la forme :

$$d^2 (S_1, S_2) = \sum_{i \in I} \alpha_i^2 (s_i^1 - s_i^2)^2$$

on peut penser que l'accroissement dû au poste  $i$  pour cette distance est  $\alpha_i^2 (s_i^1 - s_i^2)^2$ .

. si nous faisons la somme de ces parts pour tous les couples de structures de l'ensemble  $\mathcal{S}$ .

$$P_i = \sum_{S_1 \in \mathcal{S}} \sum_{S_2 \in \mathcal{S}} \alpha_i^2 (s_i^1 - s_i^2)^2$$

. si l'on effectue la somme sur  $S_2$  en conservant  $S_1$  fixe, on obtient le moment d'inertie des projections de l'ensemble sur l'axe  $i$  par rapport à la projection de  $S_1$ . Appelons alors  $G_i$  la projection du centre de gravité,  $M_i$  le moment des projections de l'ensemble par rapport à leur centre de gravité,  $N$  le nombre de structures de poids, supposé ici égal à 1, et appliquons le théorème de Huygens :

$$P_i = \sum_{S_1 \in \mathcal{S}} (M_i + N d^2 (G_i, S_1^1))$$

la somme sur tous les  $S_1$  de  $\mathcal{S}$  donne donc :

$$P_i = 2N M_i$$

de même la somme générale des distances donne :

$$P = 2N M$$

où  $M$  désigne le moment de l'ensemble par rapport à son centre de gravité. Il est donc logique de calculer l'importance relative de chacun des postes dans les distances, par la formule :

$$\text{Imp}_i = \frac{M_i}{M}$$

Les  $M_i$  sont les moments d'inertie de l'ensemble par rapport aux hyperplans perpendiculaires aux axes  $i$ , chacun des axes étant normé par  $\alpha_i$ , on a encore :

$$M_i = \alpha_i^2 \sigma_i^2$$

Choisir par exemple  $\alpha_i$  égal à  $\frac{1}{\sigma_i}$  revient alors à donner autant d'importance à chacune des variables.

Dans le cas d'une distance dans laquelle les axes des rubriques ne sont pas orthogonaux, nous ne pouvons plus appliquer cette méthode, puisque les moments par rapport aux hyperplans orthogonaux ne s'additionnent pas.

Nous avons essayé de mettre au point des systèmes de mesure des importances relatives de chacun des postes dans le cas d'une matrice non diagonale. Les deux méthodes reviennent au calcul décrit dans le cas où les rubriques sont orthogonales.

La première méthode consiste à essayer de partager le moment d'inertie totale de l'ensemble des structures en parts attribuables à chacune des variables, comme cela a été fait dans le cas d'un système de distance concernant les différentes rubriques orthogonales entre elles.

On commence par décomposer la matrice  $Q$ , comme au paragraphe 3d p. 165 en produit D.C.D. où  $D$  est une matrice diagonale de termes positifs  $\alpha_i$ ,  $C$  est une matrice comportant des 1 sur la diagonale et des termes positifs inférieurs à 1 ailleurs.

On cherche alors les vecteurs propres  $V_j$  et les valeurs propres  $\lambda_j$  de la matrice  $C$ . Après multiplication des vecteurs différence de structures par  $D$ , on les décompose sur la base  $V_j$ . On pose alors :

$$\Delta S = S_1 - S_2 = \sum_{j \in I} \xi_j V_j$$

et

$$\|\Delta S\|_Q^2 = \sum_{j \in J} \lambda_j \xi_j \|V_j\|_I^2$$

où  $\|V_j\|_I$  est la norme de  $V_j$  par la matrice unité  $I$ .

Le moment d'inertie total de l'ensemble peut alors s'écrire :

$$K = \sum_{j \in J} \lambda_j \sigma^2(\xi_j) \|V_j\|_I^2 = \sum_{j \in J} K_j^j$$

où  $K_j^j$  est le moment d'inertie du nuage projeté sur le vecteur  $V_j$ .

On décompose alors la norme de chaque vecteur du système orthogonal des  $V_j$  de la façon suivante :

$$\|V_j\|_I^2 = \sum_{i \in I} (v_j^i)^2$$

Ceci nous permet de décomposer chacun des  $V_j$  suivant des "parts" revenant à chacune des rubriques :

$$K_i^j = \lambda_j \sigma^2(\xi_j) v_j^i{}^2$$

et en sommant les  $K_j^i$  pour un  $i$  donné, on obtient :

$$K_i = \sum_{j \in J} \lambda_j \sigma^2(\xi_j) v_j^i{}^2$$

Pour une structure  $S$  donnée, on passe du système  $D.S$  (multiple de  $S$  par  $D$ ) au système  $\square$  par :

$$\square = U X$$

où  $U$  est la matrice comportant comme vecteurs colonnes les composantes de vecteurs des axes des rubriques dans la base de  $(V_j)$ .

On a donc la formule : 
$$\xi_j = \sum_i u_j^i \alpha_i x_i$$

d'où :

$$\sigma^2(\xi_j) = \sum_{i \in I} \sum_{i' \in I} u_j^i u_j^{i'} \alpha_i \alpha_{i'} \rho_i^{i'} \sigma(x_i) \sigma(x_{i'})$$

où  $\rho_i^{i'}$  est le coefficient de corrélation de  $x_i$  et de  $x_{i'}$  dans l'ensemble des structures. On peut donc calculer les  $K_i$ .

Eventuellement, si dans un système de poids ainsi calculé on veut obtenir un système donné a priori, on pourra se servir des  $\alpha_i$  comme paramètres de commande. On a des équations de la forme :

$P_i(\alpha) = K_i$  où les  $P_i(\alpha)$  sont des polynômes homogènes du second degré en  $(\alpha_i)_{i \in I}$

D'après l'exposé théorique sur la représentation de l'espace des emplois, la modification des  $\alpha_i$  peut se faire par une modification soit des distributions de probabilité attachées aux rubriques, soit de la mesure construite sur l'espace des emplois, qui, comme on l'a déjà noté, intervient dans le calcul des intégrales des carrés des fonctions densité.

La résolution d'un tel système peut se faire par une méthode d'approximation en minimisant par exemple :

$$f(\alpha) = \sum_i (P_i(\alpha) - K_i)^2$$

sur le domaine défini par les contraintes :  $\alpha_i \geq 0$   
les  $K_i$  étant tous positifs et de somme égale à 1.

Dans l'autre méthode, on considère que l'influence d'une rubrique est égale à la variation marginale que provoque sur le moment d'inertie une affinité élémentaire de même rapport sur chacun des axes.

On écrit donc :

$$K = f((\nu_i \alpha_i)_{i \in I})$$

où les  $\nu_i$  sont les coefficients d'affinité qui valent tous 1 pour le système  $(\alpha_i)_{i \in I}$ . Une affinité élémentaire de  $d\nu_i$  provoque une variation de  $K$  de :

$$\frac{\partial K}{\partial \nu_i} (1, \dots, 1) d\nu_i = \frac{\partial f}{\partial \nu_i} d\nu_i$$

Or,  $f$  est une fonction homogène du second degré en  $\nu_i \alpha_i$  puisque  $K$  l'est.

$$\begin{aligned} K &= \sum_{i \in I} \sum_{i' \in I} E(\nu_i \alpha_i \nu_{i'} \alpha_{i'} q_i^{i'}) \\ &= \sum_{i \in I} \sum_{i' \in I} \nu_i \alpha_i \nu_{i'} \alpha_{i'} q_i^{i'} p_i^{i'} \sigma(x_i) \sigma(x_{i'}) \end{aligned}$$

On pose donc :

$$K_i = \frac{\partial K}{\partial \mu_i} (1, \dots) = 2 \sum_{i' \in I} \alpha_i \sigma(x_i) q_i^{i'} p_i^{i'} \alpha_{i'} \sigma(x_{i'})$$

De plus, l'égalité d'homogénéité devient, puisque les  $i$  sont égaux à 1 :

$$\frac{1}{2} K_i = K$$

Si l'on veut ajuster les poids à un système donné, en se servant des paramètres  $(\alpha_i)$  comme commande, on doit résoudre un système d'équation du même type que dans les premiers cas.

#### 4) Les regroupements de postes

Nous allons montrer que le regroupement de deux postes peu différents apparaît comme l'atténuation progressive de leur différence, c'est-à-dire que l'on peut se rapprocher autant que l'on veut du regroupement, en conservant cependant les postes distincts.

Autrement dit, si les postes d'indice  $(n-1)$  et  $n$  sont peu différents, on peut s'arranger pour que la distance entre deux structures  $S_1$  et  $S_2$  qui ne diffèrent que par un transfert du poste  $(n-1)$  au poste  $n$ , soit aussi faible que l'on veut. Comme on l'a déjà vu, il suffit que les coefficients  $q_{n-1}^{n-1}$ ,  $q_n^n$ ,  $q_n^{n-1}$  tendent tous trois vers une même limite, puisque lorsqu'ils sont égaux la distance :

$$d^2(S_1, S_2) = q_{n-1}^{n-1} \Delta x^2 + q_n^n (-\Delta x)^2 - 2 q_n^{n-1} \Delta x (-\Delta x)$$

est nulle et que cette distance est fonction continue des trois coefficients.

Posons alors  $Q = (q_i^j)$  matrice de la forme quadratique de distance utilisée pour la nomenclature en  $n$  postes conservant les  $(n-1)$ ème et  $n-i$ ème postes distincts avec :

$$Q = \begin{pmatrix} P_{n-2} & & q_{n-1}^1 & q_n^1 \\ & & q_{n-1}^j & q_n^j \\ q_1^{n-1} & q_j^{n-1} & q_{n-1}^{n-1} & q_n^{n-1} \\ q_1^n & q_j^n & q_{n-1}^n & q_n^n \end{pmatrix} = \begin{pmatrix} P_{n-2} & & Q_{n-2}^{n-1} & Q_{n-2}^n \\ & & q_{n-1}^{n-1} & q_n^{n-1} \\ q_1^{n-1} & q_j^{n-1} & q_{n-1}^{n-1} & q_n^{n-1} \\ q_1^n & q_j^n & q_{n-1}^n & q_n^n \end{pmatrix}$$

Considérons les deux matrices

$$R = \left( \begin{array}{ccc|cc} P_{n-2} & & & \Delta_1 & \Delta_1 \\ & & & \Delta_j & \Delta_j \\ & & & \Delta_{n-2} & \Delta_{n-2} \\ \hline \Delta_1 & \Delta_j & \Delta_{n-2} & \Delta_{n-1} & \Delta_{n-1} \\ \Delta_1 & \Delta_j & \Delta_{n-2} & \Delta_{n-1} & \Delta_{n-1} \end{array} \right) = \left( \begin{array}{ccc|cc} P_{n-2} & & & S_{n-2} & S_{n-2} \\ & & & & \\ \hline {}^t S_{n-2} & & & \Delta_{n-1} & \Delta_{n-1} \\ {}^t S_{n-2} & & & \Delta_{n-1} & \Delta_{n-1} \end{array} \right)$$

$$S = \left( \begin{array}{ccc|c} & & & \Delta_1 \\ & P_{n-2} & & \Delta_j \\ & & & \Delta_{n-2} \\ \hline \Delta_1 & \Delta_j & \Delta_{n-2} & \Delta_{n-1} \end{array} \right) = \left( \begin{array}{ccc|c} P_{n-2} & & & S_{n-2} \\ & & & \\ \hline {}^t S_{n-2} & & & \Delta_{n-1} \end{array} \right)$$

Considérons un vecteur différence en postes :

$$X = \begin{pmatrix} Z_{n-2} \\ X_{n-1} \\ X_n \end{pmatrix}$$

et le vecteur Y formé à partir du vecteur X en regroupant les deux derniers postes :

$$Y = \begin{pmatrix} Z_{n-2} \\ X_{n-1} + X_n \end{pmatrix}$$

On pose alors :

$$\begin{aligned} {}^t X \cdot R \cdot X &= {}^t Z_{n-2} \cdot P_{n-2} \cdot Z_{n-2} + X_{n-1} ({}^t Z_{n-2} \cdot S_{n-2}) + X_n ({}^t S_{n-2} \cdot Z_{n-2}) + \\ & X_n (Z_{n-2} \cdot S_{n-2}) + X_n ({}^t S_{n-2} \cdot Z_{n-2}) + \Delta_{n-1} (x_{n-1}^2 + x_n^2 + 2x_{n-1}x_n) \\ &= {}^t Z_{n-2} \cdot P_{n-2} \cdot Z_{n-2} + 2(x_{n-1} + x_n) \cdot {}^t Z_{n-2} \cdot S_{n-2} + \Delta_{n-1} (x_{n-1} + x_n)^2 \\ &= {}^t Y \cdot S \cdot Y \end{aligned}$$

et :

$$\begin{aligned}
 {}^t X.R.X - {}^t X.Q.X &= {}^t X.(R-Q).X = 2x_{n-1} {}^t Z_{n-2} (Q_{n-2}^{n-1} - S_{n-2}) \\
 &\quad + 2x_n \cdot {}^t Z_{n-1} (Q_{n-1}^n - S_{n-1}) + (q_{n-1}^{n-1} - s_{n-1}) x_{n-1}^2 + (q_n^n - s_{n-1}) x_n^2 \\
 &\quad + 2 x_n x_{n-1} (q_n^{n-1} - s_{n-1}) \\
 &= {}^t Y.S.Y - {}^t X.Q.X
 \end{aligned}$$

On voit donc qu'en faisant tendre les vecteurs  $Q_{n-2}^{n-1}$  et  $Q_{n-1}^n$  coefficients rectangles des postes  $n-1$  et  $n$  avec les  $(n-2)$  premiers postes vers une même limite  $S_{n-2}$  et les coefficients  $q_{n-1}^{n-1}$ ,  $q_n^n$  et  $q_n^{n-1}$  vers  $s_{n-1}$  la distance par  $Q$  en  $n$  postes tend vers la distance par  $S$  en  $(n-1)$  postes quelles que soient les deux structures dont on mesure l'écart.

Pour conserver un système de poids fixe en regroupant progressivement les postes  $(n-1)$  et  $n$ , la distinction entre  $(n-1)$  et  $n$  devant disparaître, le vecteur imposé doit donc regrouper les rubriques  $(n-1)$  et  $n$ .

Dans le premier système les coefficients du polynome caractéristique varient d'une façon continue en fonction des coefficients  $\alpha_i$ . D'après le théorème des fonctions implicites, les racines varient alors continument dans le corps des complexes, si elles ne sont pas racines doubles. De plus, dans le cas présent, la matrice restant symétrique réelle, les racines restent dans  $R$ .

De même les vecteurs propres varient d'une façon continue. Donc les solutions en  $\alpha_i$  varient continument sous les conditions du théorème des fonctions implicites. On remarque que lorsque tous les coefficients correspondant au postes  $(n-1)$  et  $n$  sont égaux le vecteur  $\vec{e}_n - \vec{e}_{n-1}$  est vecteur propre de valeur propre nulle et que dans tous les autres vecteurs les composantes de  $\vec{e}_{n-1}$  et  $\vec{e}_n$  sont égales. Donc  $x_{n-1}$  et  $x_n$  interviennent bien dans les calculs sous la forme  $x_{n-1} + x_n$ . La restriction au sous-espaces engendré par  $(\vec{e}_1, \dots, \vec{e}_{n-2}, \vec{e}_{n-1} + \vec{e}_n)$  est donc égale à la forme quadratique sur l'espace complet. On peut donc passer à l'espace en  $(n-1)$  postes. Le regroupement de deux postes est bien un cas limite de proximité des postes aussi bien pour le



calcul des distances que des importances relatives des rubriques.

Pour le second système de calcul, on voit immédiatement que le système qui définit les  $K_i$  est à coefficients continus par rapport aux  $q_i^j$ , coefficients de la matrice servant à mesurer les distances interstructurelles. Donc si la matrice  $(q_i^j \rho_i^j \sigma(x_i) \sigma(x_i))$  qui est le jacobien du système en  $\alpha_i$ , est régulière, nous nous trouvons dans les conditions d'application du théorème des fonctions implicites et les solutions en  $\alpha_i$  sont continues. De plus, si les  $q_{n-1}^i$  et  $q_n^i$  sont égaux, deux à deux, ainsi que  $\alpha_{n-1}$  et  $\alpha_n$ , on sait que :

$$\rho_{n-1}^i \sigma(x_i) \sigma(x_{n-1}) + \rho_n^i \sigma(x_i) \sigma(x_n) = \rho_{(n-1) \cup n}^i \sigma(x_i) \sigma(x_{n-1} + x_n)$$

puisque :

$$E(x_i \cdot (x_{n-1} + x_n)) = E(x_i \cdot x_{n-1}) + E(x_i \cdot x_n)$$

Donc le système peut être écrit directement en fonction des  $(n - 1)$  postes de la nomenclature regroupée sans changer les solutions.

L'ensemble des solutions proposées est donc cohérent, il n'est certes pas unique et il est possible que l'on puisse en trouver qui soit plus facilement interprétable et qui soit plus justifié.

Il est certain toutefois que le problème de l'importance des différentes rubriques doit être considéré et qu'il ne peut se résoudre dans le cas d'une matrice non diagonale en considérant uniquement les écarts types des variables initiales, car si nous imaginons une matrice dans laquelle tous les coefficients non diagonaux soient nuls, à l'exception des coefficients rectangles concernant un certain  $i$ , alors les termes rectangles feront intervenir l'écart sur le poste  $i$ , qui de ce fait se verra attribuer une importance déterminante.

Toutefois, notre but étant d'analyser un ensemble de structures d'emploi, il serait important de poursuivre cette étude en examinant l'influence des variations des coefficients de la matrice  $Q$  sur les résultats totaux de l'analyse.

Malgré les complications que cette matrice non diagonale entraîne pour les calculs et l'interprétation, il nous semble, comme nous avons essayé de le montrer au début de cet exposé, que son introduction soit logique et nécessaire. Le calcul des distances entre structures d'emploi par une matrice diagonale, comme dans l'analyse factorielle des correspondances, n'est certainement pas neutre, puisqu'il dépend entièrement de l'arbitraire de la nomenclature utilisée.

## II L'analyse de l'ensemble des structures d'emploi

### 1) Le but de l'analyse

Chaque unité (établissement ou entreprise) est caractérisée d'une part, par les données structurelles de l'emploi, d'autre part, par les valeurs qu'elle donne à certains paramètres économiques : secteur économique, taille, localisation, chiffre d'affaires et de ventes, etc... Nous voulons détecter les liaisons qui peuvent exister entre ces deux ensembles de données. Nous avons déjà vu comment comparer les structures d'emploi. Notre problème est maintenant de regrouper des unités de structures relativement proches pour comparer leurs caractéristiques économiques aux caractéristiques de l'ensemble de la population. Nous devons donc réaliser des groupes qui rassemblent des structures semblables et qui soient assez éloignés les uns des autres. Il peut être intéressant de connaître en plus, pour chaque groupe, quelques éléments centraux (squelette du groupe) qui caractériseront le type de structures.

### 2) Rappel de quelques méthodes

Les techniques d'analyse d'un ensemble muni d'une distance sont nombreuses. Nous en retiendrons deux types : les méthodes de projection et les méthodes de partition.

Les premières utilisent la représentation de l'ensemble des points dans un espace vectoriel euclidien, ce qui est le cas ici comme nous l'avons vu précédemment. On obtient après l'application de telles méthodes une visualisation simple et fidèle de l'information.

Les secondes n'utilisent que le tableau des distances entre les objets et donnent des groupes les plus homogènes possibles.

#### . Les Projections

Dans notre étude, seules les analyses en composantes principales nous intéressent, puisque le calcul des distances entre les éléments du nuage a été déterminé plus haut et que nous avons vu l'inconvénient que présentait une analyse factorielle sur les données brutes. Le problème est donc uniquement de trouver le sous-espace de plus petite dimension absorbant la plus grande part possible de la

variance, c'est-à-dire rendant le mieux compte des distances entre les points. Cela revient à chercher l'axe le mieux ajusté, puis l'axe orthogonal au premier (au sens du produit scalaire utilisé pour la distance) le mieux ajusté et ainsi de suite. Si deux axes par exemple suffisent à recueillir un pourcentage élevé de la variance totale, la population sera représentée avec une bonne exactitude sur le plan de ces deux axes. Pour trois ou quatre axes, il faudra considérer simultanément deux plans. S'il faut plus d'axes pour avoir une perte d'information acceptable, il deviendra délicat de considérer l'ensemble des plans.

Techniquement, il sera simple de se ramener à un espace euclidien muni d'un produit scalaire de matrice unité. Il suffit pour cela de multiplier tous les vecteurs structure par une matrice qui soit racine carrée de la matrice  $Q$  utilisée pour la forme quadratique de distance. Pour interpréter ensuite les axes principaux d'inertie, il faudra revenir au repère initial en multipliant les vecteurs directeurs par la matrice inverse qui existe, puisque  $Q$  est régulier et donc ses racines carrées aussi.

#### . L'analyse hiérarchique

Pour cette analyse, seul le tableau des distances est utilisé. Le principe est de regrouper successivement les groupes les plus proches. On commence donc par fixer une distance entre groupes de points. Les plus utilisés sont, si  $d$  est la distance entre points

$A = \{x_i / i \in I\}$  et  $B = \{y_j / j \in J\}$  étant deux groupes de points :

$$d_1(A, B) = \min (d(x_i, y_j) / i \in I, j \in J)$$

$$d_2(A, B) = \max (d(x_i, y_i) / i \in I, j \in J)$$

$$D_3(A, B) = d(GA, GB)$$

où  $GA$  et  $GB$  sont les centres de gravité des groupes, s'il est possible de les déterminer. On regroupe ensuite les deux points les plus proches, puis les deux points ou groupes les plus proches et ainsi de suite jusqu'à ce que tout soit regroupé. On obtient alors une "hiérarchie" de partitions emboîtées. Nous choisirons celles qui sont les plus pertinentes pour l'analyse. On choisira dans la suite des distances séparant deux éléments regroupés, les étapes où l'accroissement est fort par rapport aux accroissements antérieurs ou postérieurs.

. Les nuées dynamiques

Cette méthode également n'utilise que le tableau des distances et donne une partition de l'ensemble en groupes, dont chacun est formé autour d'un "noyau" d'éléments centraux. De plus, en superposant les différentes partitions optimales obtenues on forme les groupes les plus stables (formes fortes) et les clivages les plus stables (frontières des formes faibles).

Nous utiliserons donc ces trois méthodes, (1), mais nous testerons également une autre méthode que nous allons exposer ci-dessous.

3) La méthode utilisée

a) Les aspects généraux

Remarquons que les méthodes de projection ne donnent pas à l'économiste qui étudie une population, des groupes, mais seulement une image correcte des situations respectives des unités les unes par rapport aux autres. S'il a ensuite besoin d'une partition de l'ensemble fondée sur des critères d'homogénéité interne des groupes et de séparation entre les groupes, il devra la faire à partir de l'image qui lui est fournie. Il procédera alors de façon empirique. Plus précisément, se trouvant en présence d'un graphe-plan sur lequel sont reportés les éléments de la population qu'il étudie (éventuellement avec des grosseurs proportionnelles à leur importance relative dans l'ensemble), il va regarder ce graphe d'assez loin pour que les points se confondent en des zones plus ou moins foncées. La teinte lui donne une image de la densité des points dans la partie considérée. Pour effectuer ensuite une classification, il ordonnera les groupes autour de zones les plus foncées, les limites de groupes passant dans les régions les plus claires. Nous essaierons donc de systématiser le processus en introduisant une formalisation de la notion de densité dans une région de l'espace et la formation des groupes s'ordonnera autour des zones de plus fortes densités. Par rapport aux méthodes

---

(1) On trouvera des exposés plus détaillés de ces méthodes dans des ouvrages spécialisés, tels que "Analyse des données multidimensionnelles" édité par le Centre d'Etudes économiques d'entreprises, "Statistiques et Informatique appliquées"

de classification qui n'utilisent que le tableau des distances, on a l'avantage de ne pas faire de groupes arbitraires en ce sens qu'une analyse hiérarchique donnera toujours des partitions sur un nuage de points issus d'une distribution normale. Nous éviterons de couper arbitrairement des distributions ne présentant qu'un seul "noyau" autour duquel se répartit l'ensemble des points du nuage.

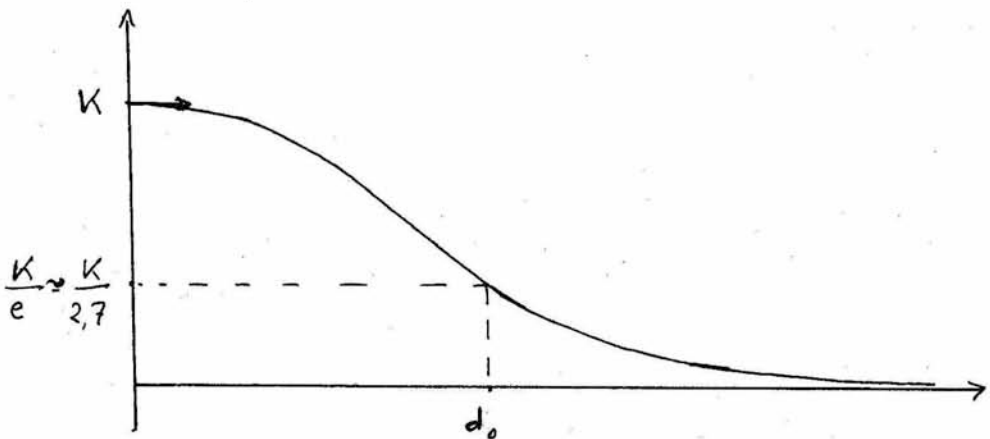
#### b) La densité

La teinte de la région regardée provient du plus ou moins grand nombre de points de cette région, mais cette donnée discontinue est lissée et donne une impression continue. Nous allons de même "répandre" chaque point en une tache foncée au centre et de plus en plus claire lorsqu'on s'en éloigne, en remplaçant par une fonction-densité de présence centrée sur l'observation qui décroitra lorsqu'on s'en éloignera. On pourrait justifier cette méthode en considérant que la donnée est entachée d'erreur aléatoire. La principale justification du procédé reste un essai de classification dans le nuage des observations. La densité totale en un point sera considérée comme la somme des densités provoquées en ce point par chacune des observations. On transformera donc une fonction discontinue en une fonction continue pour laquelle les méthodes de traitement sont beaucoup plus simples. Le problème qui se pose est celui de la forme à donner à cette fonction. Ayant déjà déterminé notre calcul de distance de façon à tenir compte des directions dans l'espace des emplois, nous devons logiquement prendre dans l'espace une fonction dont la valeur en un point ne dépende que de la distance de ce point à l'observation. La décroissance de cette fonction par rapport à la distance semble logique et nécessaire. De plus, il faudrait qu'une observation influe assez fortement et assez également sur une zone proche et très peu sur les régions éloignées. Nous devons donc prendre une fonction positive, finie au point de l'observation, faiblement décroissante d'abord, mais qui tend asymptotiquement vers zéro plus rapidement au delà d'un certain seuil. La fonction

de Gauss semble assez bien convenir à notre objet. Si on considère donc une observation A, elle crée en point X une augmentation de la densité de :

$$f_A(X) = K e^{-\frac{\|X - A\|^2}{d_0^2}}$$

où K est une constante multiplicative (égale à  $\frac{1}{(d_0 \sqrt{2\pi})^n}$  pour que l'on ait une distribution de probabilité) et  $d_0$  un paramètre. La fonction a le graphe ci-dessous.



En fait toute fonction ayant sensiblement le même graphe pourrait convenir (sous réserve théoriquement qu'elle soit intégrale). Des fonctions telles que

$$\frac{1}{1 + \frac{\|X - A\|^3}{d_0^3}} \quad \text{OU} \quad \frac{1}{1 + \frac{\|X - A\|^6}{d_0^6}}$$

ont été testées et donnent sur un nuage de 300 points les mêmes zones de forte densité. Le paramètre important est évidemment le paramètre  $d_0$ . En effet, on peut considérer approximativement que pour les points qui sont à une distance de A inférieure à  $d_0$ , l'augmentation de densité due à cette présence est sensible (supérieure à 0,3) alors qu'au-delà de  $3 \cdot d_0$  elle devient négligeable (inférieure à 0,05). La décroissance la plus rapide est aux alentours de la distance  $\frac{d_0}{\sqrt{2}}$  (point d'inflexion de la courbe).

Si l'on prend un paramètre très petit les zones de forte densité qu'on obtiendra seront celles où l'on trouvera quelques points très rapprochés relativement aux distances qui existent dans le

nuage. Ceci ne nous intéresse pas pour une analyse d'ensemble. Au contraire une valeur trop grande de  $d_0$  équivaudrait à une faible décroissance sur de très grandes régions et on noterait une densité plus forte vraisemblablement en une position centrale du nuage. Il faut donc choisir une valeur du paramètre ne tombant dans aucun de ces deux excès. On peut considérer qu'en moyenne, il doit y avoir un point donné (dans l'enveloppe convexe du nuage) un nombre  $k$  d'observations qui augmente notablement la densité. Cela nous a amené à introduire la notion de distance  $k$ -moyenne. Pour chaque point du nuage, on considère l'ensemble des distances aux autres points que l'on range en ordre croissant et on en retient la  $k^{\text{ième}}$ . On fait ensuite la moyenne arithmétique de ces distances. Cette valeur, moyenne des rayons des boules centrées sur une observation et contenant  $k$  points, tend, lorsque le nombre de points augmente indéfiniment, vers le rayon de la boule qui, centrée sur un point du nuage, contient  $k$  points. En fait, à cause des points des extrémités ou simplement proches de l'enveloppe convexe, il y a un effet amplificateur; on obtient en général une valeur telle qu'un beaucoup plus grand nombre de points sont contenus en moyenne dans ces boules. On aura donc intérêt, si l'on applique cette méthode, à diminuer les valeurs trouvées.

### c) Les zones de forte densité

Le problème de formation des groupes est avant tout un problème de repérage des zones de forte densité dans le cas où la projection sur un plan serait insuffisante. Comme nous avons remplacé chacun des points par une densité continue, les zones localement les plus denses (relativement à un environnement assez proche) correspondent maintenant à des maximums locaux de la fonction densité totale. Il nous reste à regrouper les éléments du nuage autour de ces zones de forte densité. Reprenons une image voisine de celle que nous avons déjà utilisée. Supposons que l'espace des structures d'emploi soit un plan et que nous ajoutions un axe sur lequel nous mesurons la densité en chaque point. La distribution des points structures sur l'espace de base peut alors être comparée à une distribution de masse sur un plan. Si ces masses se présentent sous une forme homogène, on obtient un relief sur le plan de base. Les zones de plus forte densité relative ou les maximums locaux sont représentés par des sommets. Supposons alors



que l'on immerge complètement ce relief et que l'on fasse alors baisser le niveau de l'eau. Au niveau donné, seule une partie du relief émerge. Le regroupement le plus logique vis à vis des considérations d'homogénéité à l'intérieur du groupe et d'hétérogénéité entre les groupes est de considérer les différentes "îles." En effet ces zones "émergeantes" sont plus denses que celles qui sont immergées, elles sont bien séparées par des zones de moindre densité. On obtiendrait ainsi une hiérarchie de partition. De plus à l'intérieur de chaque groupe, les éléments les premiers apparus sont ceux qui occupent (dans un groupe ne contenant qu'un maximum local) une position centrale. Dans un groupe contenant plusieurs maximums locaux, le "squelette" du groupe sera constitué des quelques éléments les plus proches de chacun des maximums.

d) Méthode du gradient et effet du paramètre de diffusion d

Pour appliquer la méthode, nous avons procédé de façon suivante : nous avons d'abord calculé la densité totale en chacun des points du nuage, puis nous les avons triés par ordre décroissant de densité. Partant du point de densité le plus élevé, nous suivons le gradient pour repérer le maximum local auquel il est rattaché par sa ligne de champ. Il restera ensuite pour former la hiérarchie de partition dont nous avons parlé à trouver la valeur de la densité la plus haute possible au dessus de laquelle on puisse rester en passant de façon continue d'un sommet à l'autre, valeur à laquelle se fait le regroupement des éléments rattachés à l'un ou à l'autre des sommets. Dans le cas présent, une certaine régularité des distributions nous permettra le plus souvent de prendre pour cette valeur la valeur minimale qu'atteint la densité sur le segment qui joint deux maximums locaux.

Nous avons utilisé une méthode de gradient simple sur un nuage d'environ trois cents points dans un espace à vingt dimensions. Le nombre d'itérations pour arriver au maximum local était en général assez faible. En fait, nous n'avons pas besoin d'une très grande précision quant à la localisation du maximum et nous nous sommes contentés, arrivant près du maximum local relatif à la ligne de champ d'un point donné, de tester la distance à laquelle il se trouvait des maximums déjà trouvés. Le nombre d'erreurs (définition

d'un maximum nouveau alors qu'il devrait être regroupé avec un maximum déjà trouvé) a été très faible. Un affinement du calcul pourrait certainement le réduire encore. L'inconvénient de cette application de la méthode est qu'elle prend un temps très long en ordinateur. Il est probablement possible de diminuer ce temps en utilisant par exemple un développement limité de l'exponentielle au lieu de la fonction elle-même. De même, une recherche des maximums en gradient conjugué pourrait être plus pertinente et éliminer différents risques d'erreurs. Il faut remarquer que dans les zones assez denses (1) la fonction densité totale est assez régulière et on n'assiste pas à une multiplication de maximums locaux de faible amplitude contrairement à ce que nous craignons; du moins ces maximums s'ils existent, ne sont pas repérés en raison de la grandeur du pas que nous avons choisie pour faire les calculs.

Nous avons testé la méthode sur un nuage de 370 points dans un espace à 2 dimensions, en formant les groupes selon le maximum local auquel ils se rattachaient ; il n'y a pas de résultats aberrants. Il peut arriver néanmoins qu'une scission entre deux groupes ne semble pas entièrement justifiée. Il faudrait dans ce cas pousser l'étude, voir à quelle "altitude" se situe le col séparant les deux sommets et la comparer à celle de chacun d'eux. Il est probable que ce col se situe alors assez peu au dessous des sommets.

Comme nous l'avons noté dans l'exposé général de la méthode, le paramètre  $d_0$  de diffusion influe sur les résultats de l'étude : les très grandes valeurs de  $d_0$  ne donneront qu'un seul

---

(1) Nous avons utilisé comme fonction densité  $\exp\left(\frac{-\text{IIX}-\text{AII}^2}{d_0^2}\right)$  sans coefficient multiplicatif ; la fonction densité générée par un point a alors une valeur inférieure ou égale à 1. Dans certains cas la fonction densité totale pour un nuage de 300 points atteignait un maximum de 80 environ. Jusqu'à une valeur de 6 ou 7 on aboutissait en général à des maximums qui n'étaient pas de faible amplitude puisqu'ils regroupaient un assez grand nombre de points.

maximum pour tout le nuage alors que les très petites valeurs donneront quasiment autant de maximums locaux que de points. Sur notre population plane de 370 points nous avons testé différentes valeurs de  $d_0$ . Les points qui se trouvent dans des régions relativement peu denses se trouvent assez rapidement isolés lorsqu'on abaisse la valeur du paramètre.

Voici par exemple les partitions qu'on obtient pour différentes valeurs du paramètre de diffusion :

$d_0$	Groupes de + de 10 éléments		Groupes entre 6 et 9 éléments		Groupes de 5 éléments ou moins		Nombre total de groupes	
	Nombre groupes	Nombre éléments	Nombre groupes	Nombre éléments	Nombre groupes	Nombre éléments	Nombre groupes	Nombre éléments
2.6	10	251	8	59	21	49	39	369
3.0	10	278	6	51	15	40	31	369
3.4	11	316	4	28	11	25	26	369
3.8	9	326	3	21	10	22	22	369

Ceci nous conduit à penser qu'il serait peut être bon de moduler la valeur de ce paramètre selon la situation locale de chaque point. Après avoir calculé une valeur moyenne de ce paramètre dans le nuage, par exemple par la méthode exposée plus haut, il peut être bon de lui faire subir en chaque point une homothétie dont le rapport sera supérieur à 1 si le point est dans une zone de faible densité, inférieur à 1 si le point est au contraire dans une zone très dense.

Le principal inconvénient de cette technique de calcul est sa longueur. Le temps nécessaire en ordinateur est assez grand et croît proportionnellement au carré du nombre de points. Il deviendrait très coûteux d'appliquer la méthode à un millier de points dans un espace d'une vingtaine de dimensions, du moins si l'on ne porte pas un effort sur les performances des programmes. Il nous a semblé utile de nous orienter vers une technique plus globale.

e) Une seconde technique

En fait si nous considérons que nous ne pouvons saisir tous les facteurs qui influent sur la structure d'emploi, mais seulement quelques uns qui sont prédominants, l'ensemble des structures d'emploi pourrait être subdivisé en quelques types principaux autour desquels se répartissent les structures de façon aléatoire. On peut essayer alors de chercher ces quelques distributions qui seraient les sources de la population. Ayant rendu la distribution de probabilité continue, il nous faudrait alors trouver une distribution, mélange pondéré d'un petit nombre de distributions gaussiennes, qui ajuste le mieux la densité totale calculée. On peut alors commencer par chercher la distribution gaussienne qui ajuste le mieux la fonction densité puis le couple pondéré (avec une somme de poids égaux à 1) de meilleur ajustement, puis le groupe de 3, de 4 etc...

Si le passage du groupe de  $k$  à  $(k + 1)$  distributions se fait avec une forte modification des poids et des localisations des distributions et une bonne amélioration de l'ajustement il faut itérer le procédé. Au contraire, si la  $(k + 1)$ ième étape ne se traduit que par une faible amélioration de l'ajustement, une stabilité relative des distributions déjà présentes et l'addition d'une distribution de faible poids (de l'ordre de quelques points) ou d'une distribution proche d'une distribution existante, on pourrait arrêter le procédé à l'étape antérieure.

Il nous faut décider de la forme à donner à l'appréciation de l'ajustement. Nous retrouvons ici un problème déjà abordé dans le paragraphe précédent sur l'étude des comparaisons interstructurelles, à savoir, l'établissement d'une méthode de calcul des distances entre deux distributions sur un espace métrique. L'espace de référence est ici complètement défini, c'est l'espace euclidien dans lequel nous avons représenté les structures, et la distribution de probabilité à ajuster est aussi parfaitement déterminée, puisque c'est la distribution densité totale.

C'est ici que nous demandons à nos éventuels lecteurs de se souvenir du problème tel que nous l'avons posé dans le premier paragraphe. La méthode la plus logique de mesurer cette distance est d'avoir recours à une fonction "énergie de déplacement de masse" dont il faudrait minimiser l'intégrale dans le passage d'une distribution à l'autre. Comme nous ne disposons pas de la technique nécessaire, nous en sommes réduits à nous replacer dans  $L^2(\nu)$  où  $\nu$  est ici la mesure de Lebesgue.

Néanmoins le problème n'est pas simple, car l'ensemble des sommes pondérées de la distribution gaussienne n'est évidemment pas convexe. Un barycentre de deux de ces éléments est, en général, une somme pondérée de  $2k$  distributions élémentaires. Cet ensemble est cependant fermé : une suite de distributions gaussiennes convergentes au sens de  $L^2(\nu)$  définit en effet une suite de matrices de variance-covariance convergentes au sens des distances usuelles de  $\mathbb{R}^{n^2}$ . La matrice limite de cette suite est alors la matrice de la distribution gaussienne limite de la suite des distributions. On peut donc affirmer que le problème de minimisation soulevé a des solutions.

Il serait aussi possible et peut être plus facile de prendre le problème en sens inverse, en recherchant d'abord le mélange de  $(N - 1)$  distributions gaussiennes approximant le mieux la distribution calculée pour l'ensemble de la population, puis les mélanges de  $(N - 2)$ ,  $(N - 3)$  distributions etc... Le passage d'une somme de  $(N - k)$  distributions à une somme de  $(N - k - 1)$  peut se faire en recherchant le couple des deux distributions les plus proches et en commençant par le remplacer par une seule distribution, ce qui ne semble pas très difficile.

En effet, une étude sommaire de ce problème montre qu'il y a une ou deux distributions minimisant localement dans  $L^2(\nu)$  la distance à un ensemble de deux distributions. Il suffit, dans le cas où il y en a deux, de les calculer toutes les deux et de retenir celle qui correspond au minimum absolu. Partant de ce système réduit, on ajuste l'ensemble des distributions. En fait, en utilisant d'abord le début d'une analyse hiérarchique, on peut sauter plusieurs étapes d'un seul coup en cherchant tout de suite les distributions ajustant le mieux les sommets correspondant aux groupes les plus resserrés. La méthode peut de cette manière être considérablement allégée. Chacune des optimisations successives commençant en une somme déjà assez proche de l'optimum global, on pourra ainsi pallier au moins partiellement à l'inconvénient que provoque la non-convexité du domaine des solutions admissibles.

Nous avons pensé enfin à une troisième technique :

Nous cherchons à obtenir la somme de distributions gaussiennes qui représente une distribution "macroscopique" dans le nuage, chacune des distributions rendant compte d'un certain nombre d'observations. La distribution rendant compte du plus grand nombre d'observations doit logiquement avoir son centre approximativement en le point de plus forte densité. Donc partant de l'observation pour laquelle on a la plus forte valeur de la fonction densité, on remonte au maximum absolu de la fonction densité. En ce maximum, on place une distribution qui doit ajuster le mieux possible la fonction densité dans une zone entourant ce point où l'on peut considérer qu'elle est à peu près seule à jouer, en vérifiant par exemple que les points les plus éloignés d'un groupe fixé a priori autour de ce sommet, ont bien une ligne de champ qui les mène à ce sommet. On ajuste alors sur ce groupe une fonction densité de probabilité en approximant par exemple sur un domaine ellipsoïdal autour de ce point la distribution réelle par une distribution gaussienne.

On recherche alors sur tout l'espace la zone la moins bien ajustée par cette première distribution et on itère le procédé.

On ajustera alors l'ensemble des distributions trouvées sur un ensemble de domaines. Il est possible qu'il soit préférable de ne pas considérer lors de ces ajustements tous les paramètres de chaque distribution comme variables. Ceci nous permettra en particulier de déterminer les poids de chaque distribution. On arrêtera le procédé lorsque le gain apporté par une nouvelle distribution sera faible, ou lorsque la distribution ajoutée aura un poids faible.

On aura en fin de calcul un certain nombre de distributions "macroscopiques" qui approximeront la densité totale. On pourra situer chacun des points par rapport aux différents centres de ces distributions et formé ainsi des groupes. Ces groupes devraient avoir l'avantage d'être regroupés autour de noyaux existants réellement et de ne pas être, comme cela peut se passer dans une analyse hiérarchique, le résultat de séparations dont certaines au moins peuvent paraître arbitraires.

En fait, on obtient en définitive non pas l'appartenance sure de chaque structure à un groupe donné mais des probabilités d'appartenance d'une structure donnée à chacun des groupes.

Expliquons plus précisément notre propos. Supposons l'espace des structures sur lequel auraient été ajuster  $k$  distributions macroscopiques  $D_1 \dots D_k$ . Soit alors une structure  $A$ . Les fonctions densité  $f_1, \dots, f_k$  des distributions prennent en  $A$  des valeurs  $m_1, \dots, m_k$  qui définissent pour la structure  $A$  des probabilités conditionnelles d'appartenir à l'un ou l'autre des types par les formules :

$$P_j(A) = \text{Prob}(A \in D_j) = \frac{m_j}{\sum_i m_i}$$

Ces probabilités pourraient être le point de départ d'une simulation de l'évolution, que nous essayerons plus tard d'approfondir.

*Ce document a été élaboré au Département des Synthèses par*

*Messieurs Alec CHARRAS  
et Pierre SIMULA de l'Institut d'Etudes de l'Emploi  
de Toulouse*



**CEA**

Centre d'études  
et de recherches  
sur les qualifications

58 boulevard  
92170 - VAN  
Tél : 644 - 0

**CENTRE D'ETUDES ET DE RECHERCHES  
SUR LES QUALIFICATIONS**  
(Min. Education Nationale) 58, B<sup>d</sup>. du Lycée-B.P.18-92-VANVES