

L'ANALYSE DES DONNEES AU SERVICE DE LA BIBLIOMETRIE

OUTILS DE VEILLE TECHNOLOGIQUE A LA DIMENSION DES MOYENNES ENTREPRISES

*Dou Henri, Quoniam Luc, Rostaing Hervé, Nivol William
Centre de Recherches Rétrospectives de Marseille,
Université Aix-Marseille III,
13397 Marseille CEDEX 13*

La Veille Technologique est au goût du jour. Et pour cause, le besoin en France se fait sentir. Les entreprises prennent conscience de l'indispensable nécessité de gérer l'information qui les environne. C'est à ce titre qu'un Système d'Information Scientifique et Technique, développé par une entreprise, prend toute sa fonction [1, 2].

Certains ont déjà compris que pour survivre aux "attaques" de ses concurrents il faut intégrer cette activité à celles déjà existantes. Seule de grandes entreprises ont pu se le permettre jusqu'à présent et souvent elles ont orienté ce service vers une unique gestion des portefeuilles de brevets. Actuellement les entreprises d'importance moyenne veulent aussi accéder à ces modes de surveillance et de prévision. La bibliométrie⁽¹⁾ est devenu un outil accessible à tous grâce à l'évolution des nouvelles technologies de l'ordinateur [3, 4, 5]. Le développement de traitements automatiques établis sur des méthodes d'analyse des données (ACP, AFC, Classification Automatique, AFD...) offre une nouvelle dimension au traitement de l'information.

Le travail exposé est l'illustration même du panel de traitements d'analyse des données que l'on peut mettre en oeuvre pour traduire les principales tendances d'un ensemble d'information. Ces méthodes ont permis de mener à bien une étude sur la recherche en Chimie dans les universités françaises: grouper les villes universitaires suivant leur activité scientifique. Cette étude s'inscrit fortement dans une optique Scientométrique⁽²⁾ car elle a été traitée dans un cadre Universitaire. Mais elle pourrait totalement s'appliquer à un thème ou un domaine sensible à la stratégie d'une entreprise.

(1) Système d'analyses statistiques et d'évaluation à partir de bibliographie de publications scientifiques ou de brevets.

(2) La scientométrie permet d'évaluer les divers domaines de la science et d'estimer leurs évolutions.

I. LES SOURCES : LES BASES DE DONNEES

L'utilisation et l'exploitation des bases des données est la ressource première lors de ce type d'analyse. Le gisement qu'elles constituent donne accès (par leurs exhaustivités géographiques, temporelles et thématiques) à l'information recherchée. On peut être certain, tout au moins dans le monde universitaire, que l'étude des publications que ce soit d'un pays, d'une ville ou d'un laboratoire reproduit parfaitement leurs activités scientifiques.

On peut donc bâtir notre travail à partir de cette source en toute quiétude à condition que le choix de la base ait été bien réfléchi. Elle doit avoir une bonne couverture du thème et ne doit favoriser aucun des acteurs ou des critères du phénomène à l'étude. Ce choix s'est porté sur *Chemical Abstract*⁽³⁾ qui reste incontestablement la base la plus complète et la plus exhaustive concernant l'information scientifique et technique en Chimie [6].

Pour des raisons bien compréhensibles nous nous sommes restreint qu'aux publications, pour l'année 1985, des 17 principales villes universitaires (Liste 1 p 3). Le nombre de références téléchargées sur micro-ordinateur s'est élevé à environ 6500. Le traitement d'une telle masse de données ne pourrait s'envisager sans des procédures automatiques d'analyse statistique.

Une référence est constituée de champs. Un champ représente une part de l'information globale de la référence mais est aussi, en lui même, une information. Le champ qui reproduit le mieux l'information "thème de l'article" est le champ section code de *Chemical Abstract*⁽⁴⁾ pour les raisons suivantes:

- La condensation de l'information en un code
- La non existence de termes synonymes comme dans le langage
- La pérennité dans l'espace et le temps
- Le traitement informatique rapide: gain de temps

Un tel champ est constitué d'un code qui indique le thème principal de l'article et d'une succession de sections secondaires retraçant les thèmes connexes abordés dans l'article.

A priori la simple étude des sections codes principales pourrait induire une perte d'information par rapport à une analyse plus systématique de l'ensemble des sections codes. Nous estimerons la part d'information apportée lors de l'introduction des sections secondaires pour notre exemple.

(3) Base de données interrogée sur le serveur *Orbit Search Service*. Nous remercions l'organisme *ORBIT* pour son aide lors de la réalisation de ce travail.

(4) Les thèmes de l'article sont classés parmi 80 domaines scientifiques (Liste 2 p 4). Ainsi le champ sections codes d'une référence contient autant de codes qu'il y a de thèmes connexes abordés dans l'article (thèmes principaux et secondaires).

Liste 1: Liste des 17 villes universitaires étudiées

Nom de la ville	Abréviation utilisée
Bordeaux/Talence	BOR
Clermont-Ferrand	CLE
Dijon	DIJ
Grasse	GRA
Grenoble/St Martin d'Hères	GRE
Lille/Villeueuve d'Ascq	LIL
Lyon/Villeurbanne	VIL
Marseille	MAR
Montpellier	MON
Nice	NIC
Orsay	ORS
Paris	PAR
Poitiers	POI
Rennes	REN
Strasbourg	STR
Toulon	TON
Toulouse	TOU

Liste 2: Liste des sections codes
du Chemical Abstract et de leurs significations

1	pharmacology
2	mammalian hormones
3	biochemical genetics
4	toxicology
5	agrochemical bioregulators
6	general biochemistry
7	enzymes
8	radiation biochemistry
9	biochemical methods
10	microbial biochemistry
11	plant biochemistry
12	nonmammalian biochemistry
13	mammalian biochemistry
14	mammalian pathological biochemistry
15	immunochemistry
16	fermentation and bioindustrial chemistry
17	food and feed chemistry
18	animal nutrition
19	fertilizers; solids; and plant nutrition
20	history; education; and documentation
21	general organic chemistry
22	physical organic chemistry
23	aliphatic compounds
24	alicyclic compounds
25	benzene; its derivates; and condensed benzoid compounds
26	biomolecules and their synthetic analogs
27	heterocyclic compounds(one hetero atom)
28	heterocyclic compounds(more than one hetero atom)
29	organometallic and organometalloidal compounds
30	terpenes and terpenoids
31	alkaloids
32	steroids
33	carbohydrates
34	amino acids; peptides; and proteins
35	chemistry of synthetic high polymers
36	physical properties of synthetic high polymers
37	plastics manufacture and processing
38	plastics fabrication and uses
39	synthetic elastomers and natural rubber
40	textiles
41	dyes; organic pigments; fluorescent brighteners; and photographic sensitizers
42	coating; inks and related products
43	cellulose; lignin; paper; and other wood products
44	industrial carbohydrates
45	industrials organic chemicals; leathers; fats; and waxes
46	surface-active agents and detergents
47	apparatus and plant equipment
48	unit operations and proceses
49	industrial inorganic chemicals
50	propellants and explosives
51	fossil fuels; derivatives; and related products
52	electrochemical; radiational; and terminal energy technology
53	mineralogical and geological chemistry
54	extractive metallurgy
55	ferrous metals and alloys
56	nonferrous metals and alloys
57	ceramics
58	cement; concrete; and related building materials
59	air pollution and industrial hygiene
60	waste treatment and disposal
61	water
62	essential oils and cosmetics
63	pharmaceuticals
64	pharmaceutical analysis
65	general physical chemistry
66	surface chrmistry and colloids
67	catalysis; reaction kinetics; and inorganic reaction mechanisms
68	phase equilibriums; chemical equilibriums; and solutions
69	thermodynamics; thermochemistry; and terminal properties
70	nuclear phenomena
71	nuclear technology
72	electrochemistry
73	optical; electron; and mass spectroscopy and other related properties
74	radiation chemistry; photochemistry; and photographics and other reprographic processes
75	crystallography and liquid crystals
76	electric phenomena
77	magnetic phenomena
78	inorganic chemicals and reactions
79	Inorganic analytical chemistry
80	organic analytical chemistry

II. LES METHODES MISES EN OEUVRE : L'ANALYSE DES DONNEES

Les méthodes d'analyse des données permettent de dégager, derrière une grande masse d'informations, des structures d'organisation entre ces dernières. Celles qu'on a mis en application engendrent des modèles à caractères plus descriptifs que prédictifs (méthodes situationnistes). Mais. Ces évaluations recoupées en dynamique (état des lieux à des périodes différentes) permet d'estimer les tendances majeures de l'évolution du thème à l'étude.

Les méthodes que nous avons utilisées ont toutes pour principes de replacer les objets étudiés les uns par rapport aux autres en fonction de leurs ressemblances ou leurs dissemblances. Ceci dans le but de pouvoir les regrouper suivant leur similarité de caractères. Deux approches mathématique différencient ces méthodes :

- méthodes d'analyse factorielle [7, 8]: ACP, AFC, AFD

leurs objectifs est de représenter dans un espace réduit l'information contenu dans un tableau de données. Elles renseignent, par conséquent, sur les deux entrées du tableau. Donc le but est de condenser l'information dans un espace non fini. Le critère d'optimisation de la méthode est de construire l'espace conservant la plus grande part de la diversité de l'information (variance).

- méthodes de classification automatique [9, 10, 11]:

elles ont pour but d'établir une partition parmi les individus étudiés suivant des critères de proximité et d'agrégation choisis. L'espace obtenu est un espace fini mais il ne représente qu'une des deux entrées du tableaux de données (une entrée du tableau se classe en fonction de la répartition de l'autre entrée).

III. RESULTATS

Tous les résultats présentés ont été obtenus grâce à une version améliorée par le CRRM du logiciel STAT-ITCF⁽⁵⁾.

Le tableau initial de données (Tableau 1 p 7), obtenu à partir du fichier téléchargé par traitement automatique de chaîne de caractères (logiciel DATRANS @CRRM), se constitue pour les colonnes des 17 villes Universitaires et pour les lignes des sections codes principales (suivies des secondaires pour le traitement global). Nous avons donc à traiter un tableau (matrice) à 17 colonnes et 78 (ou 158) lignes puisque 2 sections principales étaient présentes dans aucune des villes⁽⁶⁾.

La répartition des villes en colonnes et des sections en lignes nous est imposée par les analyses d'inerties qui, par leur principes mathématique, exigent une matrice comportant plus de colonnes que de lignes. Alors qu'un comportement normal consiste dans ce genre d'analyse à placer les individus étudiés sur les lignes et à distribuer les variables qui permettent de les caractériser sur les colonnes.

III.1. La classification automatique (Cf ANNEXE 1 p 18):

Connaissant le sujet de l'étude il n'est pas étonnant que la première méthode utilisée soit la classification automatique. De nombreux paramètres entre en jeu lors de l'élaboration de la stratégie classificatoire. Celle qui a été choisie est la suivante:

- classification ascendante hiérarchique
- avec calcul des distances euclidiennes
- le critère d'agrégation étant la moyenne des distances pondérées

Le critère d'agrégation offre un bon compromis entre une bonne qualité d'inertie intergroupe et intragroupe [8], ceci pour que la marginalité entre les villes ne soit pas trop influente. La méthode ascendante hiérarchique donne des résultats satisfaisant pour une matrice de taille acceptable [11]. La raison du choix de la distance euclidienne s'explique plus loin pour des raisons de compatibilité entre méthodes d'analyse⁽⁷⁾.

(5) Logiciel statistique développé par l'Institut Technique des Céréales et des Fourrages. Nous les remercions pour leur collaboration.

(6) Ces deux sections ont pour codes :

41, dyes; organic pigments; fluorescent brighteners; and photographic sensitizers

44, industrial carbohydrates

(7) La métrique euclidienne est la seule distance de proximité acceptée pour des coordonnées calculées par AFC et donc permettant une classification automatique sur les résultats obtenus par AFC

Tableau 1: Tableau des sections principales

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	STR	REN	POI	PAR	ORS	NIC	MON	LIL	GRE	CLE	TOU	TOM	VIL	ROM	NAR	DIJ	GRA
1	17.00	9.00	0.00	15.00	3.00	10.00	47.00	10.00	2.00	15.00	53.00	4.00	36.00	35.00	31.00	9.00	0.00
2	10.00	9.00	3.00	53.00	16.00	23.00	51.00	9.00	2.00	11.00	53.00	0.00	56.00	41.00	3.00	4.00	0.00
3	39.00	2.00	0.00	26.00	21.00	0.00	4.00	0.00	4.00	1.00	13.00	0.00	14.00	2.00	13.00	0.00	0.00
4	1.00	4.00	1.00	4.00	12.00	6.00	5.00	14.00	1.00	2.00	14.00	2.00	36.00	12.00	9.00	4.00	0.00
5	1.00	0.00	0.00	0.00	0.00	0.00	5.00	0.00	0.00	0.00	1.00	0.00	5.00	1.00	3.00	0.00	0.00
6	22.00	3.00	0.00	9.00	12.00	6.00	20.00	17.00	24.00	1.00	0.00	0.00	16.00	9.00	17.00	1.00	0.00
7	11.00	0.00	0.00	5.00	7.00	3.00	15.00	3.00	14.00	0.00	15.00	1.00	10.00	11.00	10.00	0.00	0.00
8	0.00	0.00	0.00	1.00	6.00	3.00	3.00	3.00	5.00	0.00	2.00	0.00	3.00	0.00	0.00	0.00	0.00
9	10.00	7.00	2.00	6.00	0.00	0.00	12.00	13.00	0.00	3.00	10.00	0.00	24.00	0.00	16.00	3.00	0.00
10	5.00	2.00	2.00	7.00	10.00	2.00	5.00	12.00	0.00	9.00	25.00	0.00	21.00	10.00	20.00	4.00	0.00
11	2.00	1.00	9.00	4.00	12.00	6.00	16.00	5.00	6.00	6.00	21.00	0.00	15.00	5.00	5.00	9.00	1.00
12	3.00	14.00	3.00	2.00	4.00	3.00	10.00	7.00	1.00	2.00	7.00	0.00	10.00	10.00	6.00	4.00	0.00
13	6.00	3.00	3.00	16.00	11.00	21.00	14.00	0.00	0.00	3.00	16.00	1.00	32.00	5.00	17.00	3.00	0.00
14	3.00	1.00	0.00	14.00	14.00	6.00	7.00	0.00	2.00	1.00	12.00	0.00	18.00	10.00	11.00	3.00	0.00
15	12.00	2.00	1.00	22.00	15.00	2.00	23.00	22.00	12.00	1.00	2.00	0.00	21.00	7.00	25.00	4.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	12.00	0.00	0.00	5.00	3.00	0.00	2.00	9.00	11.00	5.00	0.00
17	0.00	5.00	0.00	4.00	0.00	1.00	17.00	3.00	2.00	2.00	0.00	0.00	2.00	2.00	10.00	15.00	5.00
18	3.00	2.00	2.00	5.00	6.00	0.00	0.00	2.00	0.00	0.00	3.00	0.00	3.00	7.00	5.00	0.00	0.00
19	0.00	0.00	1.00	0.00	3.00	0.00	0.00	7.00	1.00	1.00	0.00	4.00	4.00	0.00	0.00	0.00	0.00
20	0.00	1.00	0.00	0.00	1.00	0.00	2.00	0.00	1.00	0.00	1.00	2.00	0.00	2.00	1.00	1.00	0.00
21	0.00	3.00	1.00	0.00	10.00	0.00	2.00	0.00	0.00	0.00	3.00	0.00	1.00	1.00	1.00	0.00	0.00
22	10.00	9.00	0.00	2.00	20.00	4.00	14.00	10.00	7.00	3.00	11.00	0.00	20.00	17.00	10.00	1.00	0.00
23	5.00	4.00	1.00	1.00	4.00	3.00	4.00	1.00	0.00	2.00	3.00	0.00	7.00	3.00	2.00	0.00	0.00
24	3.00	3.00	0.00	0.00	9.00	0.00	1.00	0.00	1.00	0.00	2.00	0.00	3.00	3.00	7.00	0.00	0.00
25	5.00	4.00	0.00	1.00	6.00	0.00	2.00	0.00	1.00	0.00	1.00	0.00	3.00	3.00	4.00	1.00	0.00
26	2.00	1.00	0.00	4.00	4.00	0.00	2.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.00	1.00	1.00
27	1.00	0.00	4.00	0.00	4.00	2.00	6.00	2.00	1.00	2.00	5.00	0.00	7.00	5.00	1.00	2.00	0.00
28	2.00	6.00	1.00	1.00	4.00	0.00	2.00	5.00	0.00	3.00	2.00	0.00	4.00	2.00	7.00	0.00	0.00
29	7.00	13.00	2.00	3.00	10.00	5.00	7.00	0.00	3.00	0.00	47.00	0.00	7.00	10.00	0.00	13.00	0.00
30	0.00	0.00	0.00	0.00	3.00	0.00	1.00	0.00	2.00	0.00	0.00	0.00	2.00	4.00	2.00	0.00	0.00
31	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00
32	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	3.00	9.00	0.00	1.00	0.00	0.00
33	0.00	0.00	0.00	1.00	0.00	0.00	1.00	3.00	16.00	1.00	7.00	0.00	7.00	0.00	1.00	0.00	0.00
34	0.00	0.00	0.00	3.00	1.00	0.00	15.00	4.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	2.00	0.00
35	3.00	0.00	0.00	3.00	1.00	1.00	7.00	2.00	7.00	5.00	1.00	0.00	3.00	0.00	1.00	0.00	0.00
36	9.00	1.00	0.00	0.00	14.00	4.00	7.00	5.00	0.00	1.00	2.00	0.00	6.00	2.00	1.00	0.00	0.00
37	1.00	0.00	0.00	1.00	0.00	0.00	2.00	1.00	1.00	1.00	12.00	0.00	10.00	1.00	0.00	0.00	0.00
38	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00
39	0.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
40	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00	3.00	0.00	1.00	0.00	0.00
41	0.00	3.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00	0.00	0.00
42	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	11.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.00
43	0.00	0.00	2.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	4.00	0.00	9.00	0.00	2.00	0.00	0.00
44	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	1.00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00
45	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	1.00	0.00	0.00
46	1.00	0.00	1.00	3.00	4.00	0.00	2.00	1.00	4.00	0.00	24.00	0.00	2.00	0.00	2.00	2.00	0.00
47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.00	0.00	6.00	0.00	0.00	0.00	0.00
48	0.00	0.00	3.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.00	0.00	3.00	1.00	0.00
49	0.00	0.00	6.00	3.00	0.00	0.00	1.00	1.00	1.00	0.00	0.00	0.00	3.00	4.00	7.00	0.00	0.00
50	1.00	3.00	2.00	0.00	5.00	0.00	14.00	3.00	15.00	0.00	0.00	0.00	6.00	4.00	7.00	2.00	0.00
51	2.00	13.00	12.00	6.00	16.00	6.00	14.00	7.00	7.00	0.00	13.00	0.00	0.00	7.00	5.00	0.00	1.00
52	0.00	0.00	0.00	1.00	0.00	1.00	1.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
53	0.00	0.00	3.00	1.00	0.00	0.00	0.00	2.00	6.00	0.00	4.00	0.00	0.00	2.00	1.00	4.00	0.00
54	4.00	2.00	12.00	0.00	14.00	0.00	2.00	3.00	30.00	1.00	7.00	0.00	10.00	3.00	10.00	4.00	0.00
55	0.00	11.00	2.00	0.00	6.00	0.00	4.00	0.00	5.00	0.00	4.00	0.00	5.00	7.00	0.00	0.00	0.00
56	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	4.00	0.00	5.00	0.00	0.00	2.00	0.00
57	0.00	1.00	0.00	3.00	1.00	0.00	1.00	2.00	5.00	0.00	3.00	0.00	4.00	1.00	1.00	0.00	0.00
58	0.00	2.00	1.00	1.00	0.00	0.00	1.00	3.00	0.00	0.00	4.00	0.00	2.00	0.00	1.00	0.00	0.00
59	0.00	10.00	5.00	5.00	0.00	0.00	4.00	0.00	3.00	1.00	7.00	0.00	0.00	11.00	10.00	0.00	0.00
60	0.00	1.00	0.00	0.00	0.00	0.00	2.00	1.00	1.00	0.00	1.00	0.00	0.00	0.00	2.00	0.00	10.00
61	4.00	2.00	0.00	1.00	0.00	1.00	16.00	0.00	0.00	5.00	0.00	0.00	10.00	3.00	6.00	1.00	0.00
62	0.00	0.00	0.00	1.00	0.00	0.00	2.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	3.00	0.00	0.00
63	7.00	3.00	3.00	6.00	66.00	4.00	4.00	5.00	43.00	1.00	11.00	0.00	20.00	13.00	10.00	2.00	0.00
64	6.00	1.00	1.00	12.00	19.00	2.00	20.00	3.00	21.00	2.00	4.00	0.00	17.00	7.00	41.00	3.00	4.00
65	0.00	0.00	11.00	7.00	5.00	0.00	5.00	5.00	1.00	0.00	0.00	0.00	46.00	10.00	7.00	0.00	0.00
66	0.00	2.00	1.00	15.00	4.00	2.00	6.00	3.00	3.00	13.00	2.00	0.00	10.00	2.00	7.00	0.00	0.00
67	2.00	0.00	0.00	1.00	5.00	0.00	1.00	1.00	11.00	20.00	1.00	0.00	3.00	0.00	9.00	1.00	0.00
68	5.00	0.00	0.00	14.00	115.00	7.00	14.00	0.00	46.00	4.00	1.00	0.00	19.00	2.00	36.00	0.00	0.00
69	1.00	2.00	5.00	5.00	26.00	0.00	1.00	0.00	32.00	1.00	2.00	0.00	10.00	0.00	0.00	0.00	4.00
70	5.00	16.00	0.00	4.00	2.00	0.00	3.00	3.00	19.00	2.00	15.00	0.00	7.00	3.00	5.00	6.00	4.00
71	10.00	14.00	3.00	13.00	190.00	15.00	22.00	37.00	45.00	3.00	28.00	1.00	57.00	34.00	37.00	7.00	1.00
72	4.00	0.00	0.00	0.00	26.00	1.00	3.00	0.00	2.00	3.00							

La comparaison de l'arbre hiérarchique de la matrice globale avec celui de la matrice des sections codes principales nous porte à croire que l'introduction des sections secondaires dans la matrice des données n'apporte que peu d'information supplémentaire (cette état de fait a été vérifié par d'autres méthodes d'analyse). Les liens entre villes ne varient pratiquement pas et la troncature effectuée (6 classes) fournit dans les deux cas la même partition (Figures ci-dessous).

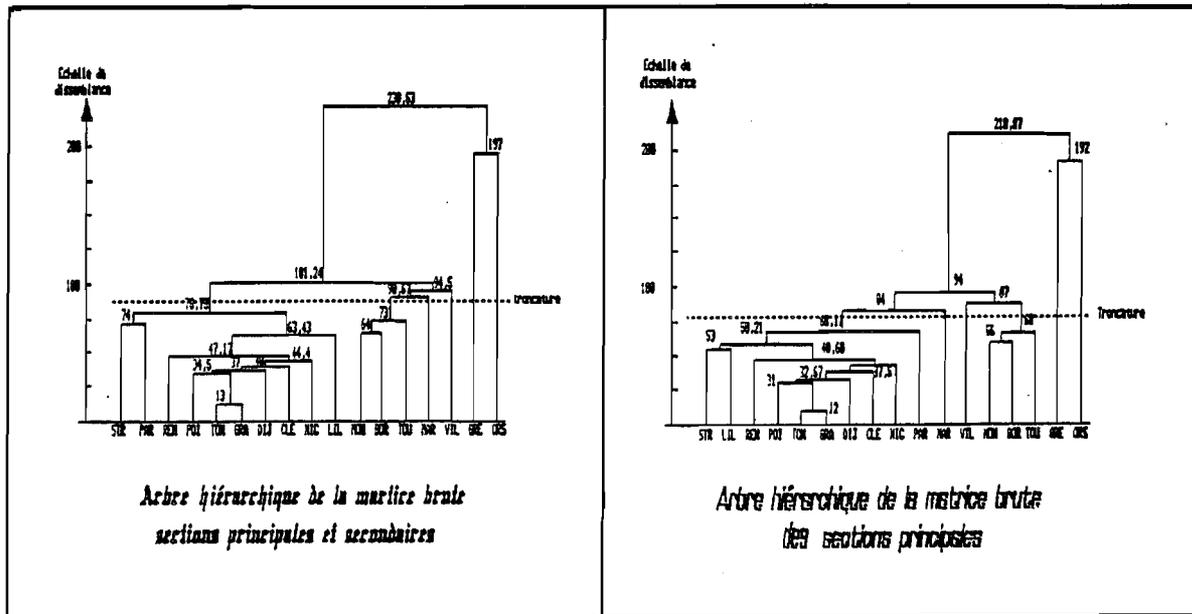


Tableau 2: Classement commun aux deux classifications (suivant les troncatures indiquées sur les figures précédentes)

Classe	Villes appartenant à la classe
1	Strasbourg, Rennes, Poitiers, Paris Nice, Lille, Clermont-Ferrand Toulon, Dijon, Grasse.
2	Montpellier, Toulouse, Bordeaux
3	Marseille
4	Lyon/Villeneuve d'ascq
5	Grenoble/S ^t Martin d'Hères
6	Orsay

Comme lors des analyses la présence d'une section secondaire dans un champ n'est pas affectée d'un poids relatif à son importance dans le champ (importance inférieure à la section principale et partagée avec les autres sections secondaires) il est probable que l'information globale en soit faussée. Il est inutile de tenter un tel risque. Donc l'analyse se poursuivra en ne considérant que les sections principales.

La méthode de classification automatique regroupe effectivement les villes suivant leurs similarités mais il est difficile d'estimer à partir de la stratégie de classification quels sont les critères de ressemblance ou de dissemblance qui sont reproduits. Il paraîtrait intéressant de savoir quelle type d'activités scientifiques se cache derrière une classe ?

III.2. Analyse Factorielle Discriminante (Cf ANNEXE 2 p 25):

Pour se faire on a essayé de savoir si une Analyse Factorielle Discriminante nous renseignerait sur cette notion. Notre espoir de réussite était faible. L'AFD permet de grouper les individus (lignes) d'une matrice entre eux ainsi que d'affecter les caractères (colonnes) aux groupes selon leur importance d'influence sur ces groupes. Mais auparavant les individus doivent avoir subi un pré-ordre.

Dans notre cas il faut donc, cette fois-ci, tout d'abord classer la matrice sur les sections codes, et non pas sur les villes, avec les mêmes principes classificatoires et la même troncature pour réaliser le pré-ordre indispensable à l'AFD. Ainsi la matrice affectée de sa nouvelle colonne peut être analysée⁽⁸⁾. Si le dépouillement des résultats livre la même partition des villes que pour la classification automatique alors les deux partitions se recouvreraient et on pourrait associer un groupe de villes à un groupe de sections codes.

Les résultats sont probants, notre exemple ne vérifie pas le cas précédemment exposé: Poitiers, Lyon/Villeneuve d'Ascq et Toulouse se retrouvent à proximité; Marseille s'est lié à Bordeaux; et Clermont est totalement isolé.

III.3. Analyse Factorielle (Cf ANNEXE 3 p 34):

La solution à tous ces désagréments est l'analyse d'inertie. Le principe mathématique est d'analyser les corrélations pour les espaces duaux⁽⁹⁾. Ces méthodes représentent les deux espaces duaux de tel façon qu'on puisse les mettre graphiquement en relation. On peut donc avoir à la fois l'information sur les individus et sur les variables. Ceci résout notre petit problème.

(8) Cette colonne établit à chaque ligne le groupe auquel appartient l'objet que représente la ligne.

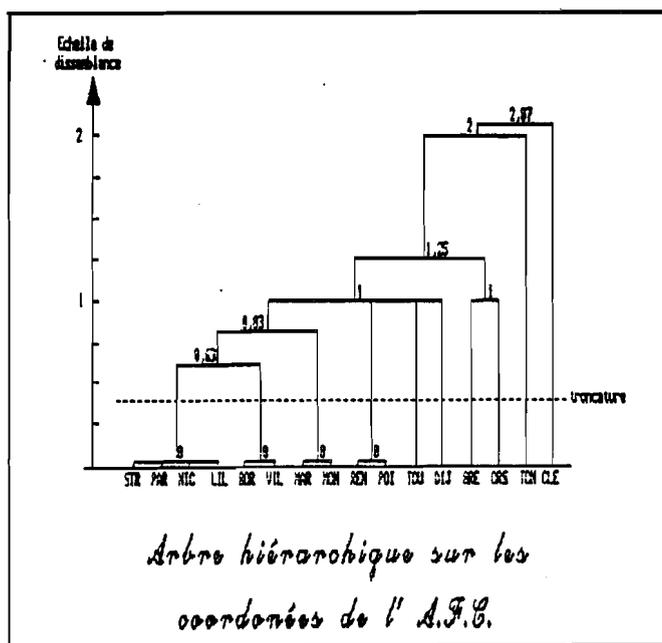
(9) La représentation vectorielle d'un tel tableau est fait soit dans l'espace des lignes soit dans l'espace des colonnes. On est donc dans l'obligation de représenter ces deux espaces (indissociables) pour connaître la totale information du tableau.

Les deux principales méthodes d'analyse factorielle utilisées en France sont l'analyse des composantes principales (ACP) et l'analyse factorielle des correspondances (AFC).

Seule la dernière méthode est applicable à notre problème. La matrice constituée a un caractère plus proche d'un tableau de contingence que d'un tableau variables-individus. Elle représente bien le croisement de deux variables définissant chacune une partition sur la population étudiée: chaque case du tableau équivaut aux références dénombrées ayant le caractère "ville" et le caractère "sections codes" commun. Ainsi dans une AFC les lignes et les colonnes jouent le même rôle. Elles sont successivement considérées comme les individus puis les variables d'une ACP, ce qui inhibe la transposition obligatoire effectuée entre les lignes et les colonnes que nous impose ces méthodes (Cf p 6).

Cette méthode mesure la ressemblance (ou dissemblance) en utilisant la distance du Khi-2, qui a l'avantage par rapport à la distance Euclidienne (de l'ACP), d'accéder à la représentation des individus des colonnes et des lignes dans un même espace. Ce qui offre l'opportunité de conserver les coordonnées des villes, en fin d'analyse d'AFC, pour enchaîner sur leur classification automatique⁽¹⁰⁾. La répartition en groupe des villes est faite automatiquement (Figure ci-dessous), ce qui n'est pas pour nous déplaire car le classement d'objets à partir des seuls résultats d'une AFC n'est pas chose aisée.

La succession des deux méthodes permet d'une part d'obtenir l'arbre hiérarchique de la classification des villes et d'autre part de retrouver les sections codes les plus proches des classes d'une troncature. La classification automatique fournit le centre de gravité de chaque classe dans l'espace de l'AFC. Il suffit de reporter ces points sur les graphes de l'AFC pour estimer les sections les plus influentes pour chacune des classes.



Il faut remarquer qu'une précaution a été prise dans l'AFC pour un meilleur résultat. La ville de Grasse a été introduite en variable supplémentaire car elle était trop marginale auprès des autres villes. Le point de cette ville placé dans l'espace réduit calculé, déformait le nuage de points et par

(10) Pour l'ACP les variables ne sont pas représentables dans l'espace cartésien où sont positionnés les objets. Seules leurs directions dans cette espace sont connues.

conséquence réduisait et distordait le champ d'information (Cf p 39). Le principal centre d'intérêt des recherches en chimie de Grasse est facilement déduit de la première AFC. On peut considérer que Grasse constitue, à elle seule, une classe qui est à ajouter à la liste de celles déterminées par la classification.

La qualité de la classification automatique après l'AFC est très bonne. A la seule vue de la hiérarchie le classement des villes est évident: la troncature sur le dendrogramme ou le regroupement sur le graphe à plat s'effectuent sans aucune hésitation (Cf p. 41 et 42). Ce qui facilite l'interprétation.

Tableau3: Récapitulatif des résultats de l'AFC
suivi de la Classification Automatique

Classe	Villes	Sections qui caractérisent la classe
1	Strasbourg, Nice Paris, Lille	2,3,6,9,13,14,15 23,26,32,40,66
2	Bordeaux, Lyon	19,22,23,25,40,65,80
3	Marseille, Montpellier	6,7,10,15,18,28 52,62,66
4	Rennes, Poitiers	20,27,29,48,57,61
5	Toulouse	12,27,48,51,62,78
6	Dijon	16,17,18,20,50 51,61,62
7	Grenoble	33,56,76,77
8	Orsay	65,70,73,74
9	Clermont	39,68,69
10	Toulon	5,20,50
10	Grasse	62

IV. COMPARAISON DES RESULTATS

Si on se penche de plus près sur les résultats des différentes analyses certains paraissent contradictoires. En fait, l'enchaînement des méthodes, comme nous venons de le faire, n'a rien de recommandable. Il a été réalisé uniquement pour démontrer que toutes ces méthodes sont applicables à des traitements bibliométriques. L'intérêt de l'analyse des données est de fournir un condensé d'information interprétable rapidement. Il ne faut donc pas retomber dans l'excès inverse en cumulant les résultats de toutes les méthodes mises à sa portée. Car même si elles sont utilisées à bon escient, elles ne feront qu'apporter des vues différentes de l'information mais pas nécessairement complémentaires. On ne peut pas recréer une information "complète" en synthétisant des "résumés" d'information établis suivant des "rhétoriques" différentes.

Pour mieux comprendre ces affirmations, nous choisissons d'exposer un exemple assez flagrant. Etudions les résultats concernant la ville de Clermont-Ferrand.

Pour l'analyse par classification automatique (ANNEXE 1 p 18), Clermont-Ferrand est positionné au sein d'un agrégat de sept villes fortement liées entre elles (Cf p 23 et 24): Rennes, Poitiers, Toulon, Grasse, Dijon, Nice et Clermont-Ferrand.

Par contre la méthode de l'AFD (ANNEX 2 p 25) marginalise la ville de Clermont-Ferrand des autres villes. Les graphes l'isolent nettement et la rattachent fortement au deux section principales 68 et 69 (Cf p 31 à 33).

La méthode de l'AFC (ANNEXE 3 p 34) livre le même résultat que l'AFD pour cette ville (p 38 et 39). La classification automatique sur les coordonnées des villes dans l'espace de l'AFC affirme encore mieux cet isolement (p 41 et 42)

Que dire de ces deux tendances apparemment contradictoires ?

C'est à la confrontation de tels résultats qu'on ressent qu'une analyse trop rapide peut déboucher sur des conclusions totalement faussées.

En fait les résultats n'ont rien de contradictoires. Il faut toujours garder en tête que chaque méthode mathématique va favoriser l'émergence de certains caractères plus que d'autres.

Ainsi la classification automatique sur les données brutes, pour le critère et la stratégie d'agrégation employés, a principalement classé les villes par l'importance de leur taux de publications. Clermont-Ferrand, étant une ville qui publie assez peu (170), est positionné dans le groupe des villes à faible taux de publications : Rennes (235), Poitiers (145), Toulon (12), Grasse(19), Dijon (228), Nice (183). Mais plus important, il est à remarquer que toutes ces villes décrivent le même phénomène: elles ont peu de publications mais elles ont toujours une domaine de la chimie concentrant la plus grande part des publications. Ce domaine prend donc une importance considérable, sauf si il correspond à un domaine à fort taux de publications nationales (dans ce cas banalisation de son émergence dans l'ensemble). Donc, bien que chacune de ces villes ait des pôles d'activités différents, la classification automatique les a regroupé pour leur phénomènes communs de marginalisation.

Inversement les deux analyses factorielles ont moins tendance à représenter les marginalités, mais plutôt de concentrer l'information pour présenter le maximum de liens entre les éléments. C'est pour cette raison que, les éléments qui ont des caractères communs se regroupent, et que plus l'élément présente l'ensemble des caractères communs à tous plus il se rapproche de l'origine du repère de l'espace. Hors Clermont-Ferrand se détache de tout nuage de points formé, ce qui indique que son activité est foncièrement différente de celle de l'ensemble des villes. Ceci s'explique par le fait que Clermont-Ferrand publie pratiquement pas dans les domaines de la chimie qui sont sources de grandes quantités de publications, tandis que vingt pour-cent de ses publications concernent les sections principales 68 et 69 où il représente vingt six pour-cent des publications nationales.

On peut donc imaginer qu'une analyse trop rapide des résultats d'une classification automatique aurait abouti à la conclusion erronée qui énoncerait Clermont-Ferrand comme faisant partie d'un groupe de villes ayant le même type d'activité scientifique.

Cet exemple permet de comprendre qu'il faut avant toute chose rechercher la méthode qui maximise les critères qu'on cherche à faire émerger de l'ensemble des données.

Notre étude n'a pas pour finalité de dégager l'innovation ou la marginalité de certains éléments mais on cherche plutôt une information organisée et faisant état de chacune des observations. Donc pour cette étude l'analyse factorielle des correspondance est celle qui répond le mieux à ces critères. La classification automatique sur les résultats de l'AFC doit alors être considérée comme une aide à l'interprétation des résultats.

L'UTILISATION DE CES METHODES DANS UNE CENTRE D'I.S.T.

Chacune de ces méthodes est à prendre comme un outil à fonctionnalité pointue. Elles sont d'un emploi facile, lorsqu'on a pris connaissance de leurs limites et de leurs domaines d'application, car toutes sont automatisées par des systèmes informatiques. Leurs analyses demandent de grandes précautions et de solides connaissances. Mais sous le flot grandissant d'information, toute personne chargée de traiter ces flux, pour en dégager les renseignements vitaux à son entreprise, est dans l'obligation de manipuler ce genre de traitements d'information au quotidien. La profession de Veille Technologique fait appel à de nombreuses compétences. La Veille Technologique devient un métier à part entière et comme toute profession elle ne peut être mise en place et exécuté efficacement que par des spécialistes. La formation de spécialiste et le développement de systèmes informatiques de traitement de l'information donneront accès très prochainement aux moyennes entreprises à cette activité d'éveille envers son environnement.

- [1] Maîtriser l'information critique
F. Jakobiak
Editions d'organisation, 1988

- [2] La veille technologique, concurrentielle et commerciale
B. Martinet, J.M. Ribault
Editions d'organisation

- [3] Veille technologique et information documentaire
De l'usage de la bibliométrie dans les services de
documentation.
H. Dou, Parina Hassanaly, L. Quoniam, Albert La Tela
Documentaliste, vol.27, n°3, mai-juin 1990

- [4] Infographic Analytical tools for decision makers
Analysis of the research production in sciences.
Application to Chemistry. Comparison between Marseille
and Montpellier.
H. Dou, P. Hassanaly, L. Quoniam
Scientometrics 17 (1989)

- [5] Teaching bibliometric Analysis and MS/DOS commands
H. Dou, Luc Quoniam and P. Hassanaly
Education for information 6 (1988) 411-423
North-Holland

- [6] Les banques de données en Chimie
L'évolution au cours des dix dernières années
A. Deroulede, C. Dutheuil
Informations Chimie n°315, mars 1990

- [7] Analyse des données (3^{ème} édition)
M. Volle
Economia

- [8] L'analyse des données
J.M. Bouroche et G. Saporta
Que sais-je, Presse Universitaires de France

- [9] Classification et analyse des données
I.C. Lerman
Dunod

- [10] Classification automatique des données
G. Celeux, E. Diday, G. Govaert, Y. Lechevallier, H.
Ralambodrainy
Dunod informatique
- [11] Algorithme de classification
M. Roux
Masson
- [12] Comment interpréter les résultats d'une analyse
factorielle Discriminante
STAT-ITCF
R. Tomassone
Institut Technique des Céréales et des Fourrages
- [13] Comment interpréter les résultats d'une analyse en
composantes principales
STAT-ITCF
G. Philipeau
Institut Technique des Céréales et des Fourrages
- [14] Comment interpréter les résultats d'une analyse
factorielle des correspondances
STAT-ITCF
R. Dervin
Institut Technique des Céréales et des Fourrages

ANNEXES 1

CLASSIFICATION AUTOMATIQUE SUR LES VILLES DE LA MATRICE GLOBALE

44

OPTIONS DEMANDEES

Classification Sur les colonnes

Classification Ascendante Hiérarchique

Distance Utilisée: Distance EUCLIDIENNE

Critère d'Aggrégation : Moyenne des Distances Pondérées

***** CLASSIFICATION AUTOMATIQUE *****

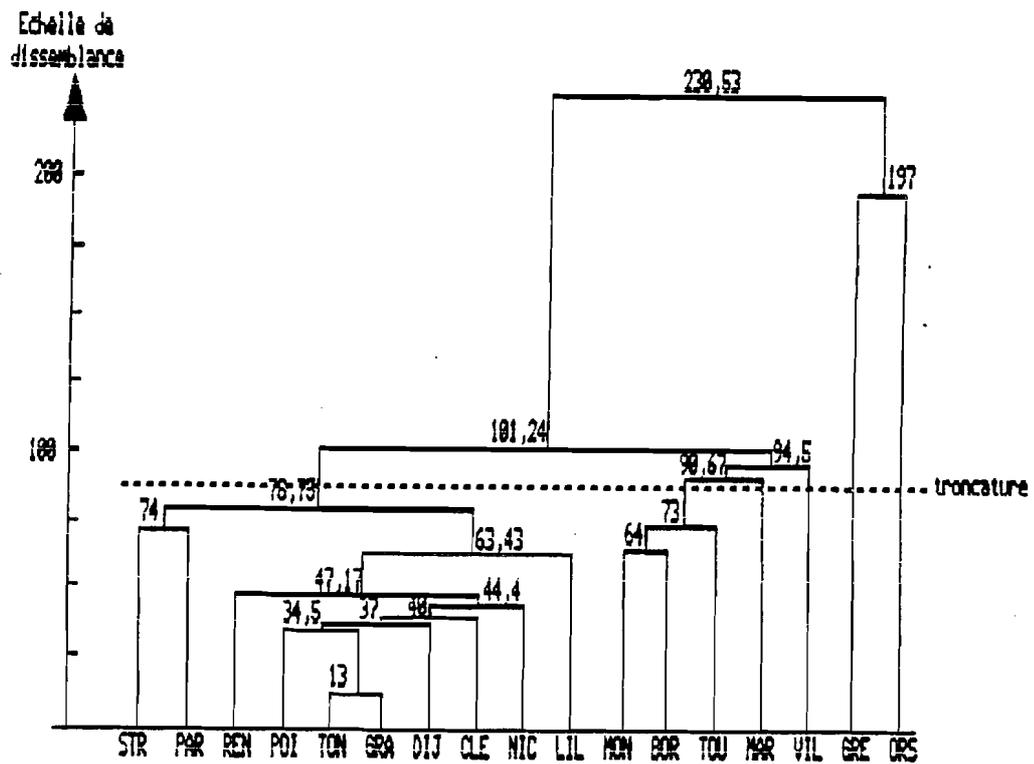
MATRICE DES DISTANCES

	STR	REN	POI	PAR	ORS	NIC	MON	LIL	GRE	CLE	TOU	TON	VIL	BOR	MAR	DIJ
REN	77															
POI	86	43														
PAR	74	77	88													
ORS	252	257	270	255												
NIC	77	49	46	56	259											
MON	90	96	100	78	240	90										
LIL	74	58	65	71	235	56	81									
GRE	201	210	219	218	197	220	195	198								
CLE	84	50	40	71	268	42	99	61	218							
TOU	107	102	119	100	242	107	78	97	202	114						
TON	91	53	33	83	279	44	117	71	220	38	131					
VIL	113	121	129	99	217	114	93	98	211	128	101	144				
BOR	86	71	88	78	227	77	64	64	185	85	68	99	89			
MAR	83	91	97	86	222	88	82	73	197	93	100	106	95	82		
DIJ	80	40	35	77	266	44	101	61	210	42	110	38	126	81	91	
GRA	92	54	36	84	279	46	118	72	220	40	133	13	146	101	100	38

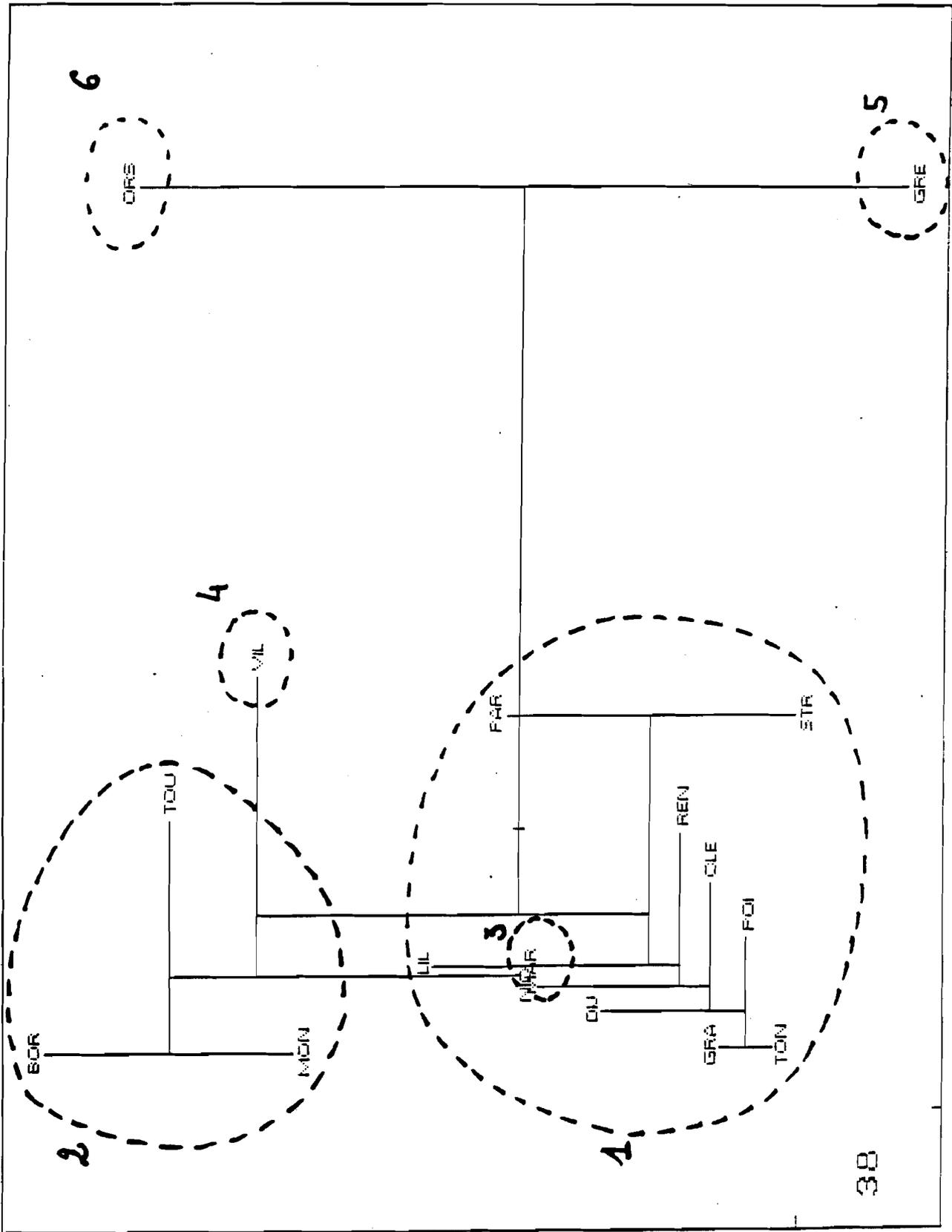
DESCRIPTION DE LA HIERARCHIE

NOEUDS	AINES	BENJAM.	POIDS	NIVEAUX
N#18	TON	GRA	2	13.00
N#19	POI	N# 18	3	34.50
N#20	N# 19	DIJ	4	37.00
N#21	N# 20	CLE	5	40.00
N#22	N# 21	NIC	6	44.40
N#23	REN	N# 22	7	48.17
N#24	N# 23	LIL	8	63.43
N#25	MON	BOR	2	64.00
N#26	N# 25	TOU	3	73.00
N#27	STR	PAR	2	74.20
N#28	N# 27	N# 24	10	78.75
N#29	N# 26	MAR	4	90.67
N#30	N# 29	VIL	5	94.50
N#31	N# 28	N# 30	15	101.24
N#32	GRE	ORS	2	197.00
N#33	N# 31	N# 32	17	230.53

ARBRE HIERARCHIQUE DES VILLES A PARTIR
DE LA MATRICE BRUTE GLOBALE



*Arbre hiérarchique de la matrice brute
sections principales et secondaires*



REPRESENTATION A PLAT DE LA HIERARCHIE PRECEDANTE

**CLASSIFICATION AUTOMATIQUE SUR LES VILLES
DE LA MATRICE DES SECTIONS PRINCIPALES**

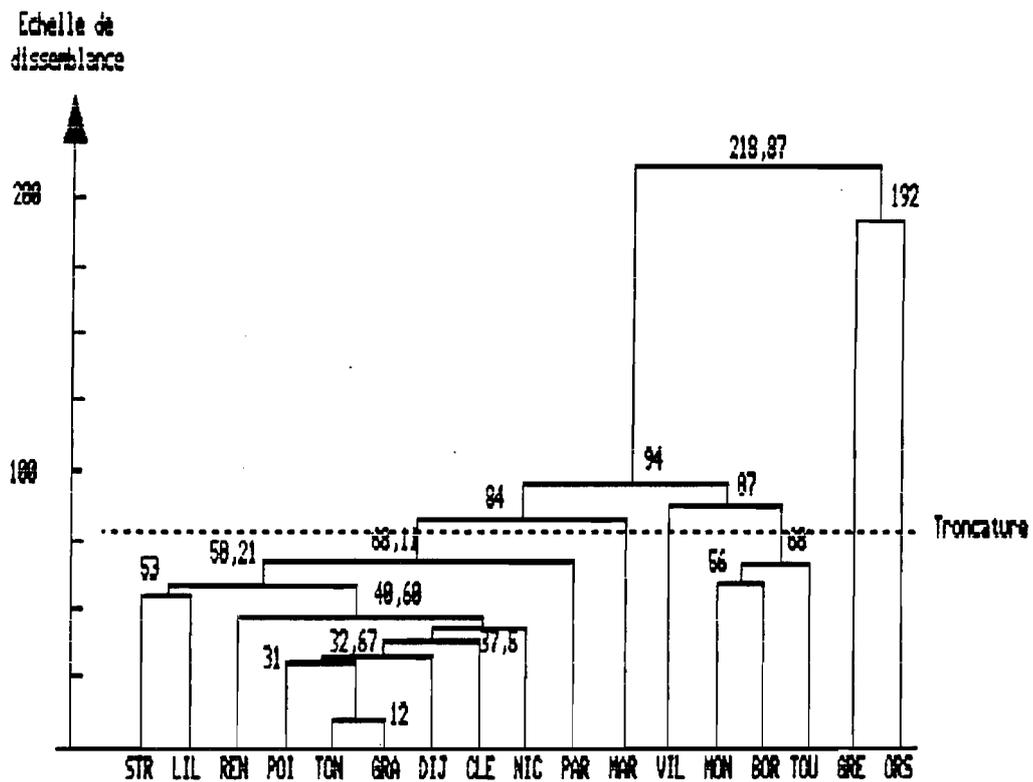
***** CLASSIFICATION AUTOMATIQUE *****

MATRICE DES DISTANCES

	STR	REN	POI	PAR	ORS	NIC	MON	LIL	GRE	CLE	TOU	TON	VIL	BOR	MAR	DIJ
REN	56															
POI	57	36														
PAR	58	71	73													
ORS	246	249	268	246												
NIC	53	43	41	53	249											
MON	82	88	96	69	234	82										
LIL	53	54	68	66	227	52	71									
GRE	188	196	281	282	192	282	181	183								
CLE	56	45	37	66	259	39	87	58	281							
TOU	99	98	111	93	237	188	73	92	198	186						
TON	68	46	38	77	267	48	185	66	218	36	122					
VIL	186	114	128	98	211	185	85	98	199	119	95	134				
BOR	73	67	81	71	221	71	56	58	171	78	63	92	81			
MAR	74	84	89	81	214	81	77	66	183	85	183	97	98	79		
DIJ	56	37	31	71	258	48	92	58	283	39	184	33	119	77	85	
GRA	62	47	32	78	268	43	186	67	218	38	124	12	136	94	98	34

	NOEUDS	AINES	BENJAM.	POIDS	NIVEAUX
N#18	TON	GRA		2	12.88
N#19	POI	N# 18		3	31.88
N#20	N# 19	DIJ		4	32.67
N#21	N# 20	CLE		5	37.58
N#22	N# 21	NIC		6	48.68
N#23	REN	N# 22		7	42.33
N#24	STR	LIL		2	53.88
N#25	MON	BOR		2	56.88
N#26	N# 24	N# 23		9	58.21
N#27	N# 25	TOU		3	68.88
N#28	N# 26	PAR		18	68.11
N#29	N# 28	MAR		11	84.88
N#30	VIL	N# 27		4	87.88
N#31	N# 29	N# 30		15	94.88
N#32	GRE	ORS		2	192.88
N#33	N# 31	N# 32		17	218.87

ARBRE HIERARCHIQUE DES VILLES A PARTIR
DE LA MATRICE DES SECTIONS PRINCIPALES



*Arbre hiérarchique de la matrice brute
des sections principales*

ANNEXES 2

**CLASSIFICATION AUTOMATIQUE SUR LES SECTIONS
DE LA MATRICE DES SECTIONS PRINCIPALES**

HIERARCHIE DECOUPEE EN 15 CLASSES

```

=====
!N°CLAS! EFFECTIF ! DESCRIPTION DES CLASSES
=====
!  1 !    43 ! P1 P2 P4 P6 P7 P9 P10 P11 P12 P13 P14 P15 P18 P19
-----
!      !      ! P21 P22 P23 P25 P27 P28 P29 P32 P37 P38 P40 P42 P45 P48
-----
!      !      ! P51 P52 P53 P57 P58 P59 P60 P61 P62 P63 P66 P67 P72 P78
-----
!      !      ! P88
-----
!  2 !    18 ! P3 P8 P24 P26 P30 P33 P36 P47 P55 P56 P65 P70 P71 P73
-----
!      !      ! P74 P75 P76 P79
-----
!  3 !     3 ! P5 P17 P20
-----
!  4 !     1 ! P16
-----
!  5 !     1 ! P31
-----
!  6 !     1 ! P34
-----
!  7 !     1 ! P35
-----
!  8 !     2 ! P39 P69
-----
!  9 !     1 ! P43
-----
! 10 !     2 ! P46 P49
-----
! 11 !     1 ! P50
-----
! 12 !     1 ! P54
-----
! 13 !     1 ! P64
-----
! 14 !     1 ! P68
-----
! 15 !     1 ! P77
=====

```

***** AIDE A L'INTERPRETATION DE LA PARTITION *****

CONTRIBUTIONS DE VARIABLES QUANTITATIVES

```

VARIANCE TOTALE      =      73
VARIANCE INTERCLASSE =     52.27
INTER/TOTAL          =     0.71

```

**ANALYSE FACTORIELLE DISCRIMINANTE DE LA MATRICE
DES SECTIONS PRINCIPALES**

NOMBRE D'OBSERVATIONS : 78 NOMBRE DE VARIABLES : 18

NOMBRE DE GROUPES : 10
NOMBRE DE VARIABLES QUANTITATIVES : 17

NOMBRE D'AXES DEMANDES : 5

ETUDE PAR GROUPE

GROUPE	EFFECTIF	VARIABLES	MOYENNES	ECARTS-TYPES DES SERIES
1 (1)	43	STR	4.116	5.863
		REN	4.870	4.627
		POI	2.442	3.142
		PAR	5.870	8.940
		ORS	6.289	6.810
		NIC	2.884	4.947
		MON	8.558	18.986
		LIL	4.698	5.471
		GRE	4.558	5.986
		CLE	1.977	3.246
		TOU	18.721	12.829
		TON	8.186	8.691
		VIL	11.860	12.891
		BOR	6.791	8.351
		MAR	7.558	8.994
DIJ	2.442	3.194		
GRA	8.279	1.515		
2 (2)	18	STR	5.444	7.967
		REN	2.722	4.628
		POI	2.800	3.266
		PAR	5.111	6.624
		ORS	37.800	48.431
		NIC	2.111	3.695
		MON	6.500	9.580
		LIL	5.722	9.700
		GRE	24.444	31.807
		CLE	1.611	2.831
		TOU	7.889	11.885
		TON	0.856	0.229
		VIL	11.222	12.634
		BOR	7.444	11.931

MAR	8.111	11.175
DIJ	1.667	2.683
GRA	8.111	8.316

3 (3) 5

STR	0.288	0.488
REN	1.288	1.939
POI	0.888	0.888
PAR	1.488	1.744
ORS	8.488	8.498
NIC	0.288	0.488
MON	18.288	5.776
LIL	1.488	1.744
GRE	0.888	0.748
CLE	1.488	1.968
TOU	2.888	2.713
TON	8.488	8.888
VIL	1.888	1.833
BOR	3.888	3.833
MAR	5.888	4.684
DIJ	6.288	5.836
GRA	1.888	2.888

4 (4) 1

STR	0.888	8.888
REN	1.888	8.888
POI	8.888	8.888
PAR	8.888	8.888
ORS	8.888	8.888
NIC	8.888	8.888
MON	8.888	8.888
LIL	8.888	8.888
GRE	8.888	8.888
CLE	4.888	8.888
TOU	1.888	8.888
TON	8.888	8.888
VIL	2.888	8.888
BOR	8.888	8.888
MAR	8.888	8.888
DIJ	8.888	8.888
GRA	8.888	8.888

5 (5) 1

STR	3.888	8.888
REN	8.888	8.888
POI	8.888	8.888
PAR	3.888	8.888
ORS	1.888	8.888
NIC	1.888	8.888
MON	7.888	8.888
LIL	2.888	8.888
GRE	7.888	8.888
CLE	5.888	8.888
TOU	1.888	8.888
TON	8.888	8.888
VIL	3.888	8.888
BOR	8.888	8.888
MAR	1.888	8.888
DIJ	8.888	8.888
GRA	8.888	8.888

6 (6) 3

STR	2.888	1.633
REN	8.667	8.943
POI	8.333	8.471
PAR	5.333	6.848

ORS	3.333	1.700
NIC	0.667	0.943
MON	2.667	2.337
LIL	1.333	1.247
GRE	4.667	4.643
CLE	11.333	7.846
TOU	1.000	0.816
TON	0.000	0.000
VIL	4.667	3.859
BOR	0.667	0.943
MAR	5.333	3.859
DIJ	0.333	0.471
GRA	0.000	0.000

54

7 (7) 2

STR	7.000	7.000
REN	0.500	0.500
POI	0.000	0.000
PAR	1.000	0.000
ORS	21.000	21.000
NIC	1.000	1.000
MON	4.500	4.500
LIL	0.000	0.000
GRE	61.000	50.000
CLE	0.500	0.500
TOU	2.000	1.000
TON	0.000	0.000
VIL	1.000	1.000
BOR	7.500	6.500
MAR	2.500	1.500
DIJ	0.000	0.000
GRA	0.000	0.000

8 (8) 2

STR	0.000	0.000
REN	0.000	0.000
POI	0.000	0.000
PAR	0.000	0.000
ORS	0.000	0.000
NIC	0.500	0.500
MON	0.000	0.000
LIL	0.000	0.000
GRE	0.500	0.500
CLE	0.000	0.000
TOU	2.000	0.000
TON	0.000	0.000
VIL	0.000	0.000
BOR	0.000	0.000
MAR	0.000	0.000
DIJ	0.000	0.000
GRA	0.000	0.000

9 (9) 1

STR	0.000	0.000
REN	0.000	0.000
POI	3.000	0.000
PAR	0.000	0.000
ORS	0.000	0.000
NIC	0.000	0.000
MON	0.000	0.000
LIL	1.000	0.000
GRE	0.000	0.000
CLE	0.000	0.000
TOU	0.000	0.000
TON	1.000	0.000

		VIL	0.000	0.000
		BOR	0.000	0.000
		MAR	3.000	0.000
		DIJ	1.000	0.000
		GRA	0.000	0.000
10 (10)	2	STR	0.000	0.000
		REN	0.000	0.000
		POI	0.000	0.000
		PAR	1.000	0.000
		ORS	0.000	0.000
		NIC	0.500	0.500
		MON	1.500	0.500
		LIL	1.500	0.500
		GRE	0.000	0.000
		CLE	0.000	0.000
		TOU	0.000	0.000
		TON	0.000	0.000
		VIL	0.500	0.500
		BOR	0.000	0.000
		MAR	1.500	1.500
		DIJ	0.000	0.000
		GRA	0.000	0.000

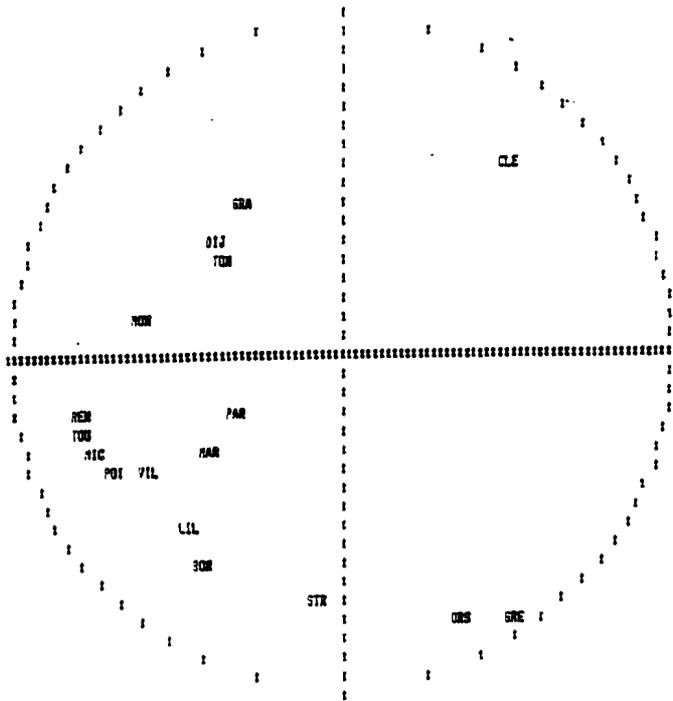
Axe	Valeur propre	Inertie	Pseudo F	WILKS	ddl	Proba	Corrél
1	1.4867	37.7%	11.23	174.40	153	11.35	0.5979
2	0.9943	25.2%	7.51	116.55	129	75.69	0.4986
3	0.9114	23.1%	6.89	72.72	105	99.30	0.4768
4	0.3159	8.0%	2.39	31.58	84	100.00	0.2401
5	0.1427	3.6%	1.08	14.15	65	0.00	0.1249

ETUDE DES CENTRES DE GRAVITE DES GROUPES

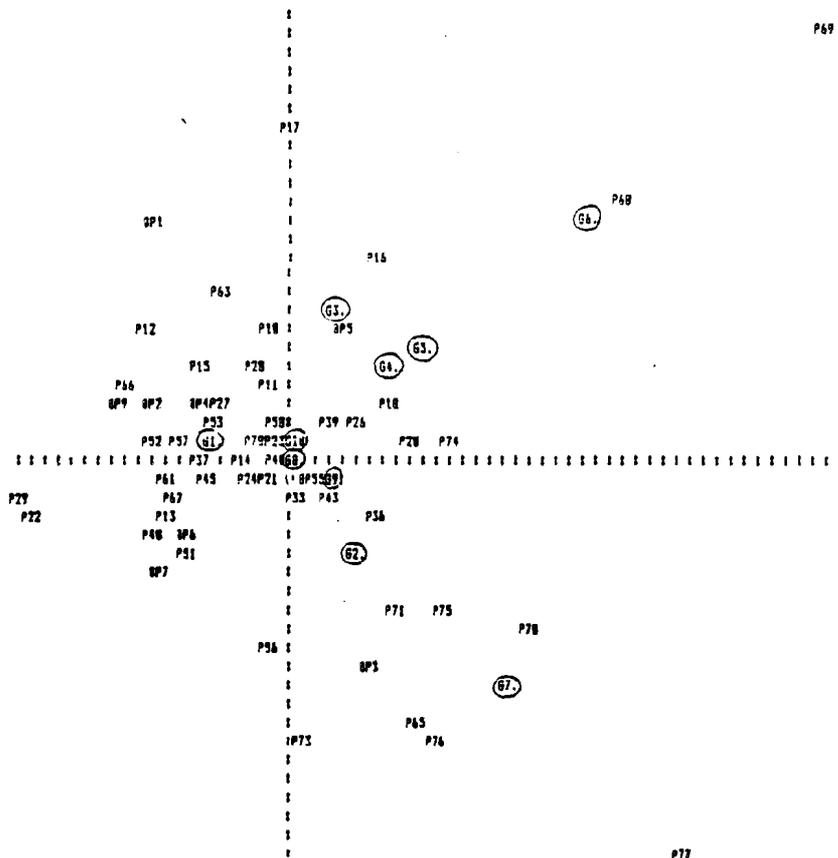
Pour chaque AXE :
 1RE COLONNE : COORDONNEES DES INDIVIDUS SUR LES AXES DISCRIMINANTS
 2E COLONNE : COSINUS CARRES (QUALITE DE LA REPRESENTATION)

GROUPE	AXE 1	AXE 2	AXE 3	AXE 4	AXE 5
1(1)	-0.5884	0.8163	0.1217	0.0349	0.2072
2(2)	0.5554	0.2327	-0.8328	0.5232	-0.0433
3(3)	0.4073	0.8245	1.2392	0.2268	-2.2323
4(4)	0.8283	0.2867	0.7382	0.2277	0.6107
5(5)	1.0692	0.3277	0.8201	0.1928	0.1472
6(6)	2.4013	0.4740	1.8899	0.2936	1.6635
7(7)	1.7387	0.2392	-1.7468	0.2414	-0.5666
8(8)	0.0778	0.0839	-0.0797	0.0841	-0.0794
9(9)	0.4105	0.8213	-0.1688	0.0836	-0.9698
10(10)	0.0625	0.0023	0.0771	0.0835	-0.2282

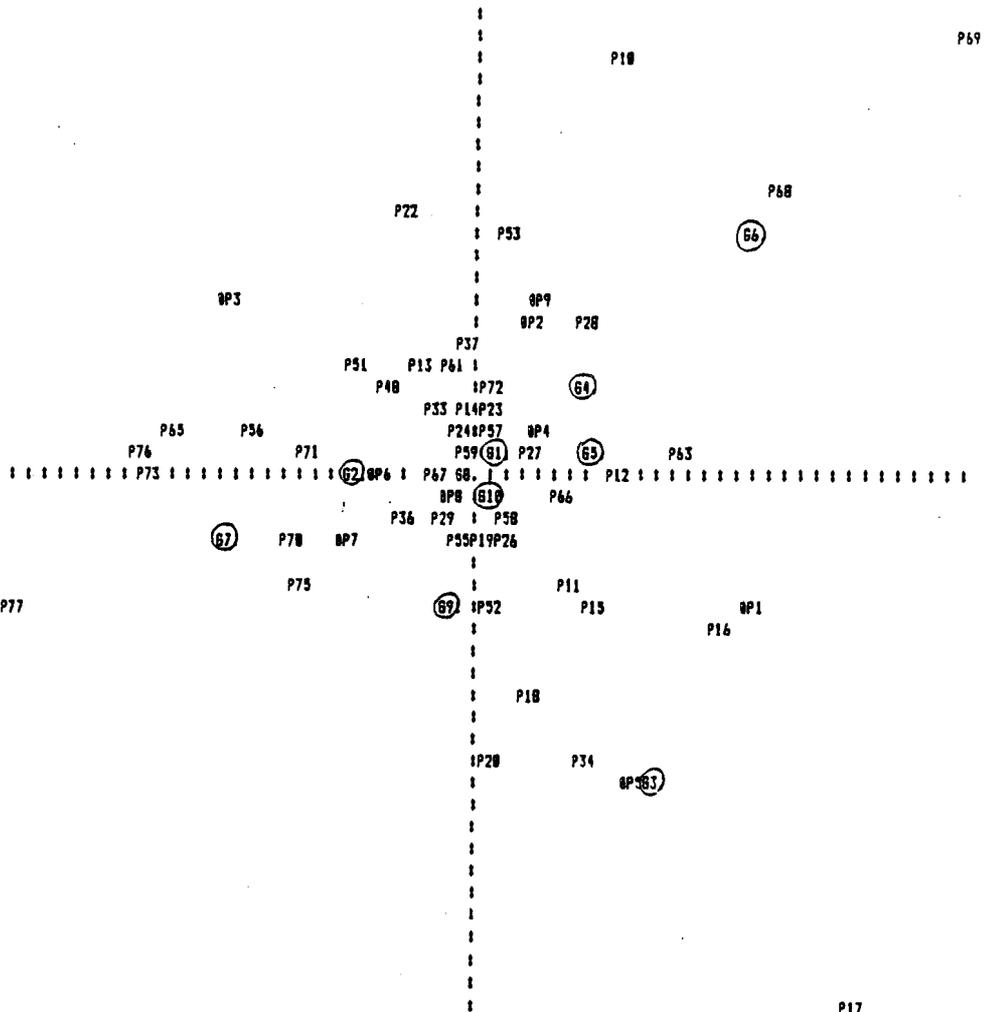
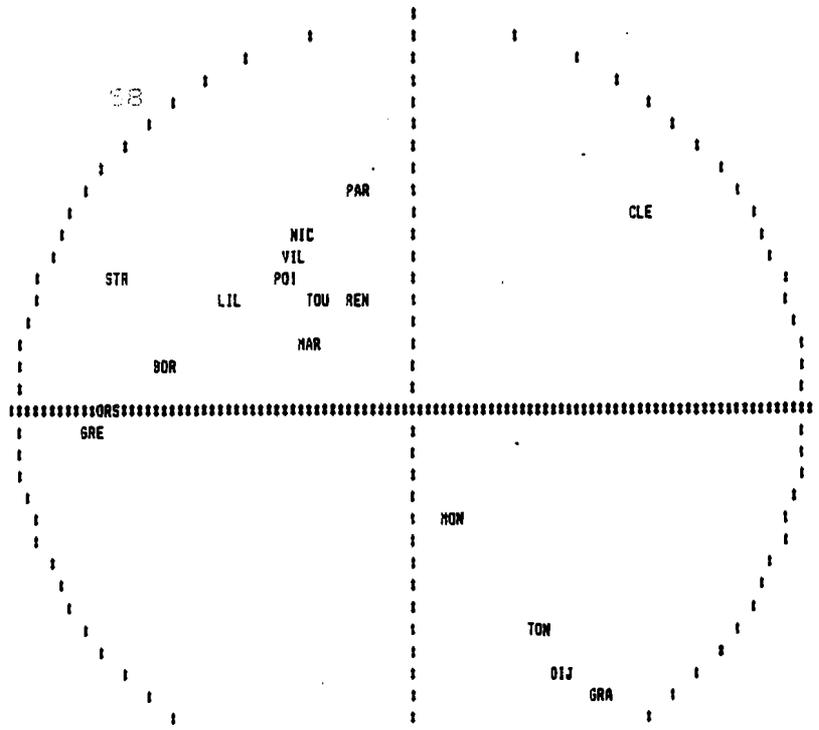
PLAN : 2 AXE 1 HORIZONTAL AXE 2 VERTICAL



PLAN : 2 AXE 1 HORIZONTAL AXE 2 VERTICAL



- | | | | |
|------------|-----|---------------|-----|
| POINT VU : | P14 | POINT CACHE : | P19 |
| POINT VU : | 60. | POINT CACHE : | P25 |
| POINT VU : | 8P0 | POINT CACHE : | P30 |
| POINT VU : | 64. | POINT CACHE : | P31 |
| POINT VU : | 8P0 | POINT CACHE : | P32 |
| POINT VU : | P20 | POINT CACHE : | P34 |
| POINT VU : | 65. | POINT CACHE : | P35 |
| POINT VU : | 60. | POINT CACHE : | P38 |
| POINT VU : | 60. | POINT CACHE : | P42 |
| POINT VU : | 8P0 | POINT CACHE : | P46 |
| POINT VU : | 60. | POINT CACHE : | P47 |
| POINT VU : | 60. | POINT CACHE : | P49 |
| POINT VU : | 69. | POINT CACHE : | P50 |
| POINT VU : | 60. | POINT CACHE : | P54 |
| POINT VU : | 60. | POINT CACHE : | P59 |
| POINT VU : | P14 | POINT CACHE : | P66 |
| POINT VU : | P37 | POINT CACHE : | P62 |
| POINT VU : | 610 | POINT CACHE : | P64 |
| POINT VU : | P52 | POINT CACHE : | P72 |
| POINT VU : | P53 | POINT CACHE : | P79 |
| POINT VU : | P21 | POINT CACHE : | P80 |



- POINT VU : 0P8
 - POINT VU : 010
 - POINT VU : 0P8
 - POINT VU : 04.
 - POINT VU : 08.
 - POINT VU : 05.
 - POINT VU : 010
 - POINT VU : 010
 - POINT VU : 08.
 - POINT VU : 010
 - POINT VU : 0P8
 - POINT VU : 023
 - POINT VU : 08.
 - POINT VU : 010
 - POINT VU : 08.
 - POINT VU : 09.
 - POINT VU : 010
 - POINT VU : 0P7
 - POINT VU : 08.
 - POINT VU : 010
 - POINT VU : 0P8
 - POINT VU : 01.
 - POINT VU : 067
 - POINT VU : 08.
- POINT CACHE : P21
 - POINT CACHE : P25
 - POINT CACHE : P30
 - POINT CACHE : P31
 - POINT CACHE : P32
 - POINT CACHE : P35 P77
 - POINT CACHE : P38
 - POINT CACHE : P39
 - POINT CACHE : P40
 - POINT CACHE : P42
 - POINT CACHE : P43
 - POINT CACHE : P45
 - POINT CACHE : P46
 - POINT CACHE : P47
 - POINT CACHE : P49
 - POINT CACHE : P50
 - POINT CACHE : P54
 - POINT CACHE : P60
 - POINT CACHE : P62
 - POINT CACHE : P64
 - POINT CACHE : P74
 - POINT CACHE : P78
 - POINT CACHE : P79
 - POINT CACHE : P80

ANNEXE 3

51

**ANALYSE FACTORIELLE DES CORRESPONDANCES DE LA MATRICE DES
SECTIONS PRINCIPALES AVEC GRASSE EN VARIABLE SUPPLEMENTAIRE**

NOMBRE DE VARIABLES (Colonnes) ACTIVES DU TABLEAU : 16
 NOMBRE DE VARIABLES (Colonnes) SUPPLEMENTAIRES : 1

NOMBRE D'AXES DEMANDES : 5

VALEURS PROPRES ET VECTEURS PROPRES

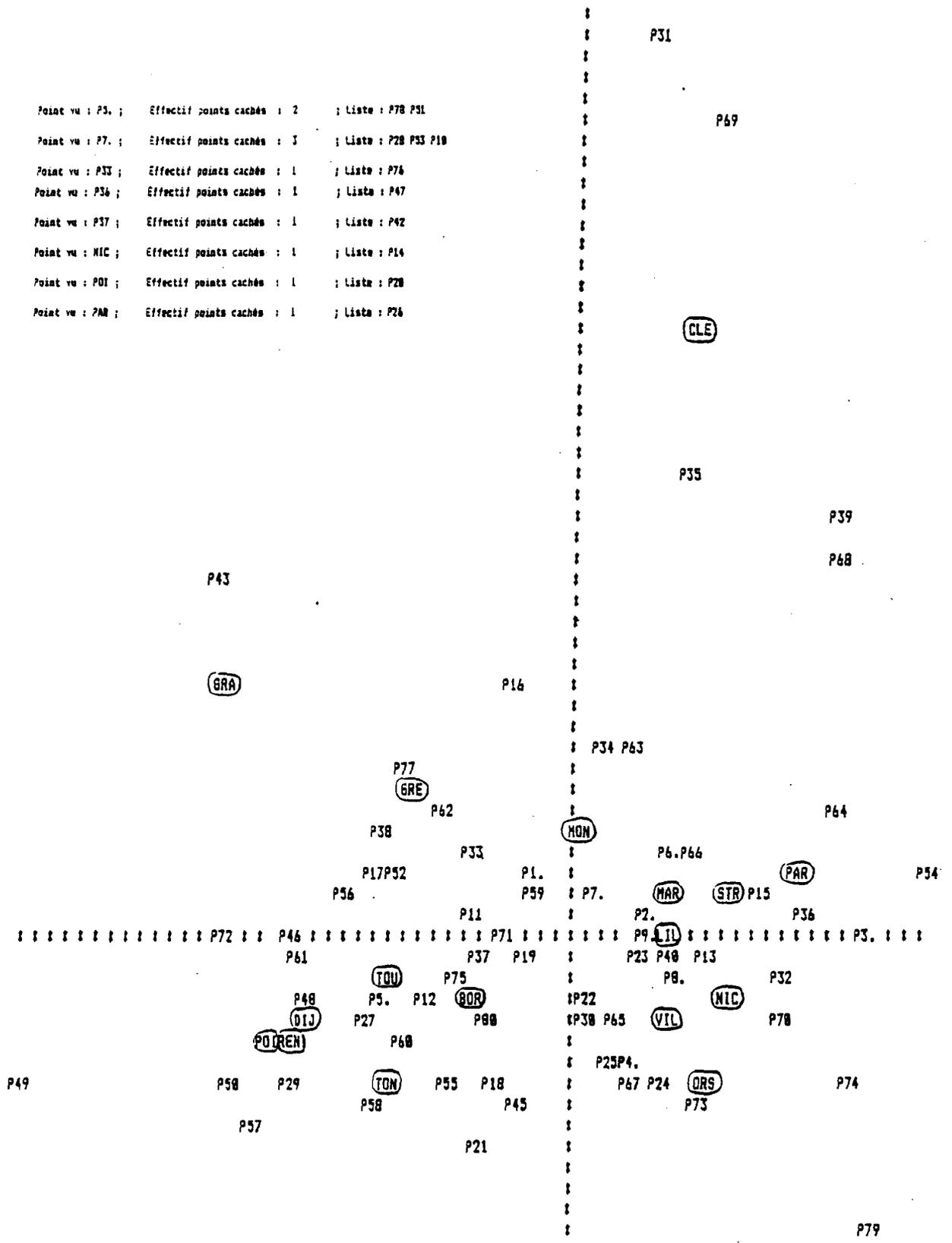
1ère LIGNE : VALEURS PROPRES (VARIANCES SUR LES AXES PRINCIPAUX)
 2ème LIGNE : CONTRIBUTION A L'INERTIE TOTALE (POURCENTAGES EXPLIQUES PAR LES AXES PRINCIPAUX)

0.2350	0.1878	0.0908	0.0738	0.0638
25.7 %	11.8 %	9.9 %	8.8 %	6.9 %

VECTEURS PROPRES (COEFFICIENTS DES VARIABLES DANS L'EQUATION LINEAIRE DES AXES PRINCIPAUX)

STR	0.1899	0.9968	0.2922	1.3772	-0.3857
REN	0.7865	-1.8714	-0.8344	-1.7827	-0.8128
POI	0.5729	-2.8841	-0.7788	-1.6744	0.5375
PAR	0.7251	1.4736	0.4834	1.2129	-0.6214
ORS	-1.3189	0.8666	-1.8927	-0.9259	-0.1783
NIC	0.7265	1.8215	-0.4643	0.5828	-1.1964
MON	0.6928	0.8243	0.7675	0.7288	0.9697
LIL	0.4118	0.6831	-0.8128	0.4873	0.3445
GRE	-1.9248	-1.8753	1.8282	0.8323	0.8874
CLE	0.7157	0.7613	4.4811	-3.6565	-1.4355
TOU	0.8288	-1.1977	-0.3895	0.6439	-1.5784
TOM	1.7888	-1.2863	-1.8388	-1.8661	6.6834
VIL	0.5934	0.6881	-0.5689	-0.8936	-0.2323
BOR	0.3495	-0.6422	-0.4162	-0.8529	0.8186
NAR	0.2915	0.5928	0.3852	-0.3691	1.7635
DIJ	1.2234	-1.7256	-0.6178	-0.4276	2.7323

REPRESENTATION SIMULTANEE DES LIGNES (Observations) ET COLONNES (Variables) \$\$\$
 PLAN 2 3 AXE 2 HORIZONTAL AXE 3 VERTICAL



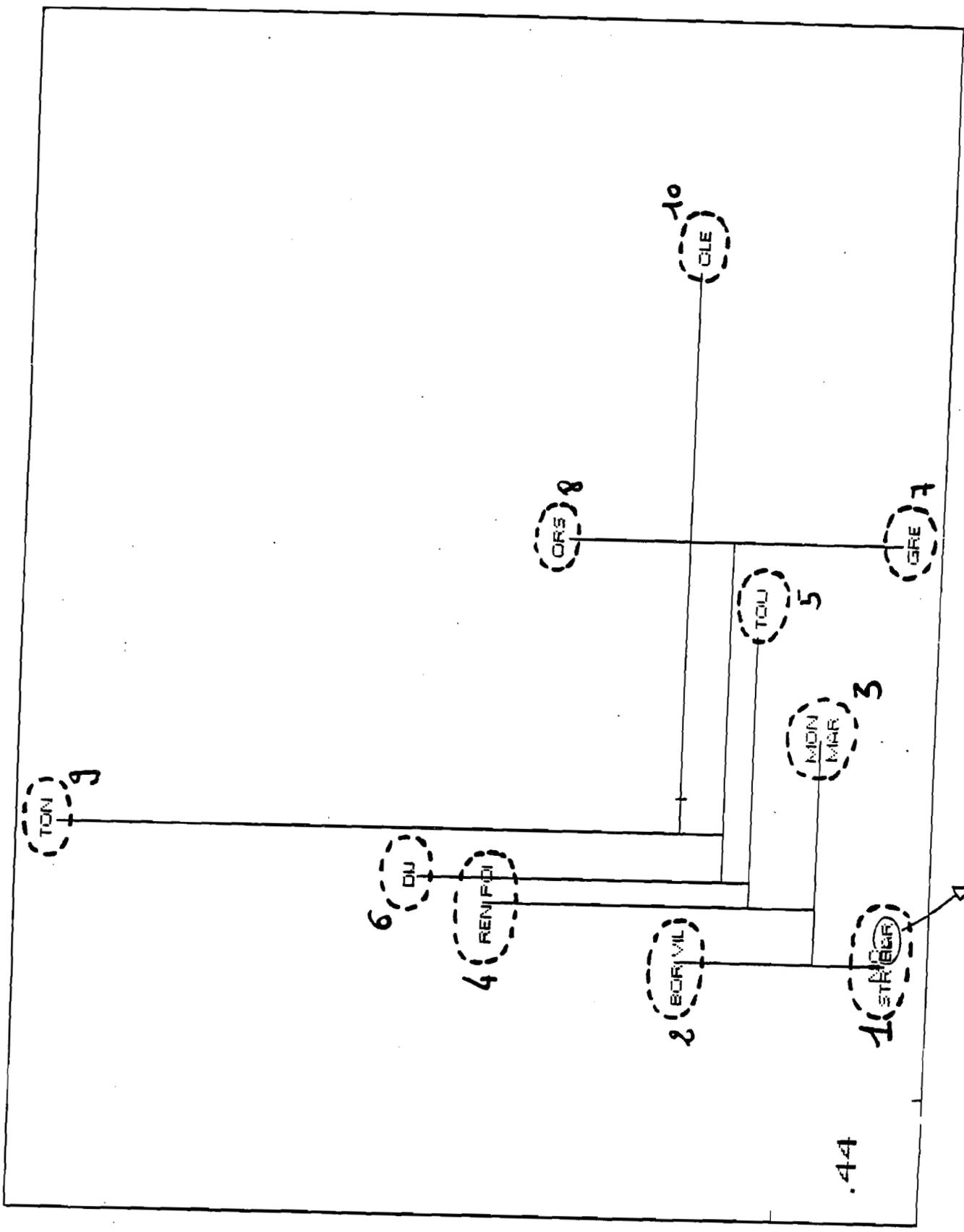
CLASSIFICATION AUTOMATIQUE SUR LES COORDONNEES
DES VILLES CALCULES PAR L' AFC PRECEDANTE

***** CLASSIFICATION AUTOMATIQUE *****

MATRICE DES DISTANCES

	STR	REN	POI	PAR	ORS	NIC	MDN	LIL	GRE	CLE	TOU	TON	VIL	BOR	MAR
REN	1														
POI	1	0													
PAR	0	1	1												
ORS	1	1	1	1											
NIC	0	1	1	0	1										
MDN	1	1	1	1	1	1									
LIL	0	1	1	0	1	0	0								
GRE	1	2	2	2	1	2	1	1							
CLE	2	2	2	2	2	2	2	2	2						
TOU	1	1	1	1	1	1	1	1	1	2					
TON	2	2	2	2	2	2	2	2	3	3	2				
VIL	1	1	1	1	1	0	1	0	1	2	1	2			
BOR	1	1	1	1	1	1	1	0	1	2	1	2	0		
MAR	1	1	1	1	1	1	0	0	1	2	1	2	1	1	
DIJ	1	1	1	1	2	1	1	1	2	2	1	1	1	1	1

	NOEUDS	AINES	BENJAM.	POIDS	NIVEAUX
N#17	REN	POI		2	0.00
N#18	STR	PAR		2	0.00
N#19	N# 18	NIC		3	0.00
N#20	MAR	MDN		2	0.00
N#21	BOR	VIL		2	0.00
N#22	N# 19	LIL		4	0.00
N#23	N# 22	N# 21		6	0.63
N#24	N# 23	N# 20		8	0.83
N#25	N# 24	N# 17		10	1.00
N#26	N# 25	TOU		11	1.00
N#27	N# 26	DIJ		12	1.00
N#28	GRE	ORS		2	1.00
N#29	N# 27	N# 28		14	1.25
N#30	N# 29	TON		15	2.00
N#31	N# 30	CLE		16	2.07



.44

PAR
LIL

REPRESENTATION A PLAT DE LA HIERARCHIE PRECEDANTE