

SAGOMBAYE NODJIRAM

M.-M. THOMASSIN

**Analyse du vocabulaire des titres des publications
du comité français de cartographie**

Les cahiers de l'analyse des données, tome 17, n° 4 (1992),
p. 471-480

http://www.numdam.org/item?id=CAD_1992__17_4_471_0

© Les cahiers de l'analyse des données, Dunod, 1992, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ANALYSE DU VOCABULAIRE DES TITRES DES PUBLICATIONS DU COMITÉ FRANÇAIS DE CARTOGRAPHIE

[TITR. C.F.C.]

*SAGOMBAYE NODJIRAM**

*M.-M. THOMASSIN***

1 Structure des données analysées

Dans un précédent article (cf. [S.M.F. – LIOUVILLE], in *CAD*, Vol. XVII, n°2, 1992), l'un des auteurs (S. N.) a montré le parti qu'on pouvait tirer de l'analyse, sur une longue période, du vocabulaire mathématique des titres des articles de deux périodiques, le Bulletin de la S.M.F. et le Journal de LIOUVILLE; chaque table annuelle étant considérée comme un fragment de texte, et l'ensemble de ces fragments, croisé avec le vocabulaire retenu.

L'étude proposée ici n'a pas la même ampleur: il s'agit seulement de quelque 800 titres de publications récentes des membres du Comité Français de Cartographie relevées par le deuxième auteur (M.-M. Th.) et rangées en 29 blocs thématiques; lesquels correspondent aux divisions et subdivisions usuelles de la nomenclature - à des fusions près rendues nécessaires par la rareté des publications comprises sous certaines rubriques.

Comme dans l'étude précédente, on a d'abord construit une liste des mots (ou, plus exactement, des formes de mots: le singulier étant, e.g., compté séparément du pluriel); on a éliminé les mots outil: articles, conjonctions, prépositions... ainsi que des mots pleins, manifestement dépourvus de tout contenu technique se rapportant à la cartographie; et on a retenu un lexique de 144 formes, qui comptent, chacune, dans le corpus des titres, au moins 5 occurrences (à l'exception du singulier "zone" qui n'en a que 4). Ce tri a laissé subsister des mots sans intérêt thématique: par exemple, si "Varsovie" figure 5 fois, c'est seulement comme le siège d'un congrès international... Mais comme ces mots n'ont pas perturbé l'analyse, on les a conservés.

(*) Étudiant en Doctorat, à l'Université Pierre et Marie CURIE.

(**) Géographe à l'ORSTOM.

Hist	Histoire de la cartographie
Docu	Problèmes généraux de documentation, d'archivage, de diffusion
Droi	Droit d'utilisation, de reproduction, d'édition, etc...
OrgF	Organisation - nationale - France
OrgE	Organisation internationale - Étranger
Ensg	Enseignement cartographique et formation de spécialistes
Défi	Définition de la Cartographie - classification des cartes
Coll	Collections de cartes
Util	Utilisation des cartes
Topn	Toponymie et transcription
Géné	Cartographie générale de base
Topg	Cartes topographiques
Aéro	Cartes aéronautiques
Mari	Cartes marines
Cada	Cartes cadastrales
Atls	Atlas - manuels
Stat	Inventaires thématiques: Enquêtes, sondages, statistiques
Natu	Environnement naturel
Sol	Occupation du sol (zones industrielles, vignobles, voies...)
Écnm	Démographie, Économie
Tech	Rédaction cartographique, projections, observations
Grph	Expression graphique: couleurs, symboles, tracé, lettres
Autm	Automatisation de la cartographie: généralités
Banq	Banques de données
InfG	Infographie
Numé	Cartographie numérique
Satl	Cartographie à partir d'images prises par satellites
Prog	Matériels - Logiciels
Impr	Matières premières des cartes et reproduction photomécanique

Liste des 29 blocs thématiques, avec leurs sigles

Le tableau ci-dessus explique le contenu des 29 blocs thématiques, avec leurs sigles. Chacune des 144 formes que compte le lexique a, de même, reçu un sigle; formé, dans la plupart des cas, de ses 4 premières lettres, à des modifications près introduites pour éviter toute confusion: e.g., {geog, geoq, Geoq} sont les sigles des trois formes {géographie, géographique, géographiques}. (Notons, au passage, que, pour le traitement des accents, on a procédé comme dans [SMF-LIOUVILLE], en utilisant le programme 'trigalac'). Mais comme, dans la suite, les formes figurent *in extenso* dans le tableau de la classification et n'apparaissent sur les graphiques d'analyse factorielle que par des numéros de classe, il n'est pas utile d'en énumérer les sigles.

En variant le découpage des blocs thématiques, on a fait plusieurs analyses factorielles; jusqu'à obtenir des résultats que les blocs les plus légers ne perturbent pas; la CAH montrant entre les blocs des proximités qui s'accordent avec la structure hiérarchique de la nomenclature de départ.

Dans la publication des résultats, on a choisi de placer d'abord la CAH; puisque c'est d'après une partition des formes en 22 classes que le nuage des mots apparaît sur les graphiques plans en correspondance avec l'ensemble des 29 blocs thématiques; mais on fera d'emblée des allusions aux résultats de l'analyse factorielle.

Hist		40	43	49		53		54		55	56	.
OrgÈ												
Mari												
Droi		35		47								
Topn												
OrgF		31		41	44							
Topg												
Écnm			37									
Défi		32										
Util												
Natu				39								
Cada			38									
Aéro		33										
Coll												
Docu						52						
Géné		34		45	46		50	51				
Tech												
Numé												
Grph												
Stat				42								
Prog		30										
Autm												
Sol				48								
Banq												
Atls												
Ensg												
InfG		36										
Satl												
Impr												

2 Classification des blocs thématiques et classification des formes de mots

Au sommet de la hiérarchie des blocs, se sépare Impr, où l'on traite de la reproduction des cartes et des propriétés mécaniques de leurs supports (papier ou plastique); puis, la classe 36, {InfG, SatL}, dominée par les techniques nouvelles d'acquisition et de traitement de l'information; puis Ensg, enseignement spécialisé dans les universités; et, enfin Atls, qui se rapporte, notamment aux utilisateurs des écoles et des lycées.

Reste alors la classe 53, qui comprend les 24 autres blocs, groupés en deux classes: 49 et 52. Comme on peut le vérifier en relisant les titres, ou d'après la correspondance avec les classes de mots, la classe 49 se rapporte principalement à l'utilisation des cartes telle qu'on la conçoit depuis plusieurs décennies; et rentre dans les grandes divisions "Études générales" ou "Cartographie: généralités".

Dans 52, on trouve Docu, avec des généralités sur l'archivage; puis, dans 51, avec ses principales subdivisions {46, 42, 48}, une conception des fonctions de la carte et de ses moyens d'expression qui, sans être toute récente, ne remonte pas très loin dans le temps. Le lecteur qui ne dispose pas de la liste de base des titres de publications, jugera du contenu de ceux-ci d'après la distribution des classes de mots, donnée ci-après en un tableau.

c Partition en 22 classes : Sigles des mots de la classe numéro c	
1	aéronautique
257	essai classification normalisation
264	analyse présentation généralités problèmes représentation exemple aide
	environnement végétal végétation technique réalisation carte F3--- F5--
250	régions nouvelle échelle zone urbaines
260	française national quelques perspectives rapport états France
	topographiques plan territoire
261	problème rôle cartes communication aménagement inventaire Tokyo cas
	utilisation
234	nouvelles état relief siècle historique
242	littoral techniques satellites photographies graphique
266	évolution cartographiques cartographique cartographie conférence
	internationale recherche son Japon développement monde géographique
	Michelin organisation tendances institut F1--
254	traitement systèmes automatique assistée ordinateur automatisation F9--
202	thématique statistiques thématiques
253	marines aspects F9+++
251	couleurs emploi procédé composition mise couleur révision F2++ F6++ F8--
245	nationale bibliographie plans documentation F7--- F5+++
265	économique Varsovie régionaux population scolaires atlas F4---- F3----
212	école Paris géographie pédagogique enseignement formation
	Université cartographes F1--- F3+++
240	agricoles centre zones occupation sol
222	cartographe cadre géographiques
262	topographique résumé base systèmes information méthode région milieu
	étude application données F1+++
248	synthèse numériques numérique F8-- F9++
259	reflexion types production data apport télédétection image images
	landsat imagery mapping satellite thematic spot F1++++
241	reproduction impression plastiques stabilité dimensionnelle F2+++++

Ci-dessus, le tableau de la partition des formes de mots est donné avec un étiquetage sommaire en terme de facteurs: il apparaît que la structure ne s'explique adéquatement qu'en recourant à des facteurs de rang élevé, ce qui (cf. *infra*) correspond à la lente décroissance des taux d'inertie afférents aux axes successifs. On verra, sur les plans (1, 2) à (1, 5), la disposition du nuage des centres de classes.

Un étiquetage clair et interprétable est fourni par le tableau de contingence $t_{iq \times jq}$, croisant classes de mots et classes de thèmes, créé en même temps que le listage VACOR d'aide à l'interprétation de la CAH. On lit, e.g., à l'intersection de la ligne i240 et de la colonne j48, qu'il y a 20 occurrences des formes de la classe i240, {agricoles,...}, dans les blocs {Sol, Banq} de la classe j48.

mots du lexique x blocs thématiques

22012	j43	j35	j41	j39	j2	j46	j42	j48	j16	j6	j36	j29
i1	0	0	2	3	0	0	0	0	0	0	0	0
i257	1 +	6	5	2	0	2	1	0	2	0	4	4
i264	15	0	20 +	44	3	11	14	17	13	3	9	6
i250	0	0	4 +	12	0	1	2	4	1	2	3	0
i260	10	1 +	33	13	5	6	14	4	1	12	0	0
i261	17	3 +	51	13	11	12	12	8	3	20	12	4
i234 +	16	0	0	1	2	5	2	0	0	0	1	0
i242	1	0	1	0	3 +	14	4	2	0	1	7	1
i266 +	80	8	61	29	12	47	51	14	5	74	21	10
i254	1	0	0	0	3	5 +	33	3	0	6	6	2
i202	2	0	3	0	1	0 +	22	1	1	5	1	0
i253 -	6	0	1	0	0	4	0	0	0	0	0	0
i251	1	0	1	1	0 +	23	4	1	0	1	3 -	9
i245	2	1	3	0 +	19	0	0	1	2	0	2	1
i265	4	1	14	3	4	0	4	1 +	31	3	0	0
i212	1	0	3	1	1	1	4	1	2 +	103	1	0
i240	0	0	0	1	1	1	0 +	20	2	0	9	0
i222	1 +	7	0	0	2	0	3	6	0	5	2	1
i262	4	0	11	5	3	10	16 +	41	0	4	34	6
i248	0	0	0	1	0 +	10	0	2	0	1	4	0
i259	2	0	5	0	1	14	2	4	2	7 +	95	0
i241	1	1	0	2	1	0	0	0	0	0	0 +	32

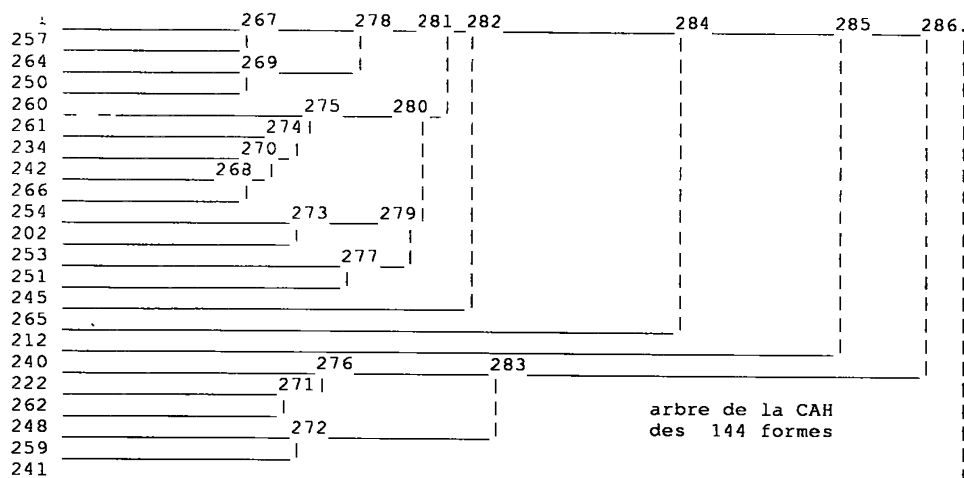
Ci-dessus: *croisement entre classes de mots et classes de thèmes*Ci-dessous: *rappel de la classification des blocs thématiques*

Hist	OrgÉ	Mari	43	49	53	54	55	56	
OrgF	Topg	Écng	Droï	Topn	35	47			
			Défi	Util	41	44			
Natu	Cada	Aéro	Coll	39					
			Docu	2		52			
Géné	Tech	Numé	Grph	46	50				
	Stat	Prog	Autm	42					
		Sol	Banq	48					
			Atls	16					
			Ensg	6					
	InfG	Satl	36						
		Impr	29						

Un signe '+' dans ce tableau indique une forte contribution, à la fois de la ligne au profil de la colonne et réciproquement; le signe '-' (horizontal) marque les cas où il y a seulement contribution de la colonne au profil de la ligne. La plupart des classes thématiques apparaissent ainsi caractérisées par une ou deux classes de formes.

La caractérisation de {Impr} par i241 est très nette. Avec {InfG, Satl}, va i259, vocabulaire où l'on trouve quelques mots en anglais qui, par le sens, s'accordent avec le reste de la classe. Mais on notera que la forme {satellites}, au pluriel, n'est pas dans i259 mais dans i242, {littoral, photographies,...}; et caractérise j46, ensemble de thèmes liés aux fonctions de la carte et à ses moyens.

On voit sans surprise dans i212 le vocabulaire de {Ensg}; et dans i265, (avec 'Varsovie', siège d'un congrès, cf. *supra*) celui de {Atls}.



Dans la CAH des formes, se sépare, au plus haut niveau, la classe 241 caractéristique de Impr: la proximité entre 241 et Impr apparaît clairement sur le plan (1, 2), présenté ci-contre.

Puis se sépare la classe 283, dont on a considéré plusieurs subdivisions, et qui comprend les caractères propres à {InfG, Satl}, {Sol, Banq}, spécialités nouvelles; ainsi qu'à {Droi, Tpnm}, dans 222: mais il n'y a, en fait, dans cette petite classe, que des formes d'un usage commun: {cartographe, cadre, géographiques}.

Les classes 212 et 265, qui se séparent ensuite nettement, et ne se subdivisent qu'à un niveau très bas (i.e. inférieur à celui des 21 nœuds retenus pour définir la partition considérée ici), sont respectivement associées à Ensg et Atls: blocs dont la CAH des thèmes montre aussi l'isolement.

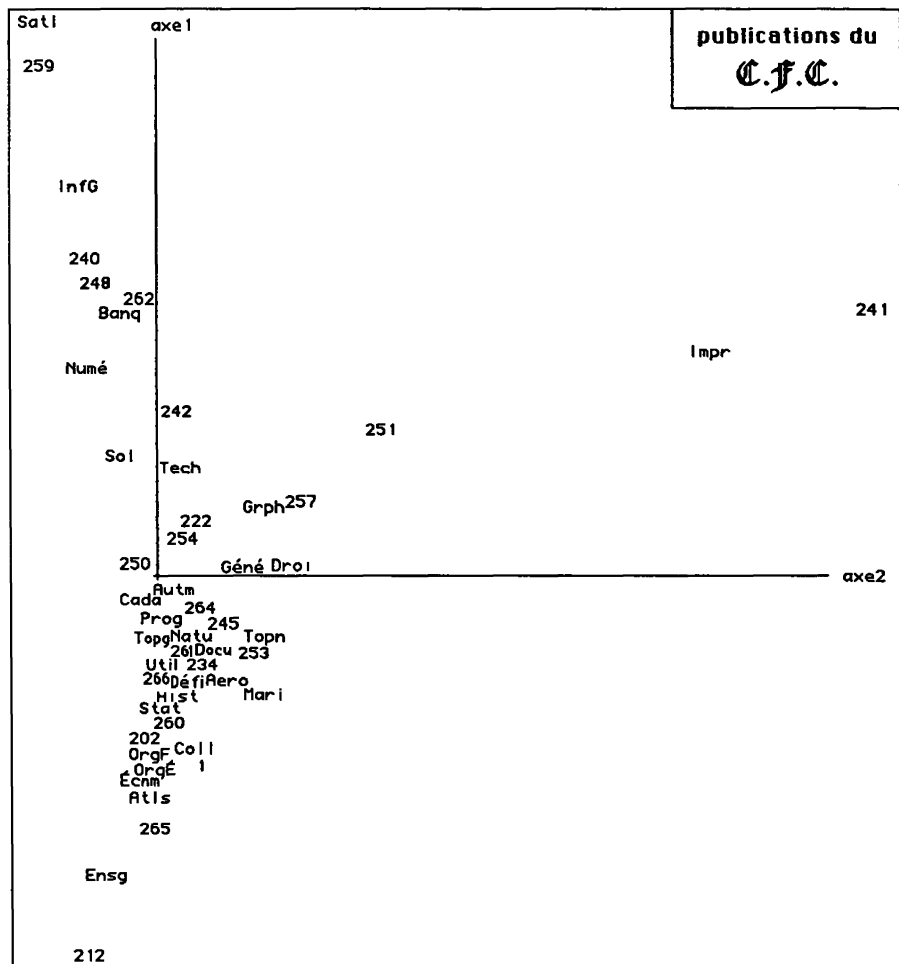
À un niveau moins élevé, se sépare encore, isolément, la classe 246, { nationale, bibliographie, plans, documentation }, caractéristique de Docu.

Reste la classe 281, dont les ramifications se distribuent de façon assez équilibrée.

Dans 279, on a les caractères lexicaux de {Géné, Tech, Numé, Grph} et {Stat, Prog, Autm}, eux-mêmes agrégés dans la CAH des thèmes pour constituer la classe j50.

La classe 275 reforme principalement les caractères de {Hist OrgÉ Mari} et {OrgF, Topog, Écnm Défi Util}.

Enfin dans 278, on remarque les caractères de {Natu, Cada, Aéro, Coll}: ainsi définis, Aéro et Cada, bien que légers, reçoivent une place satisfaisante relativement à l'ensemble des blocs de thèmes.



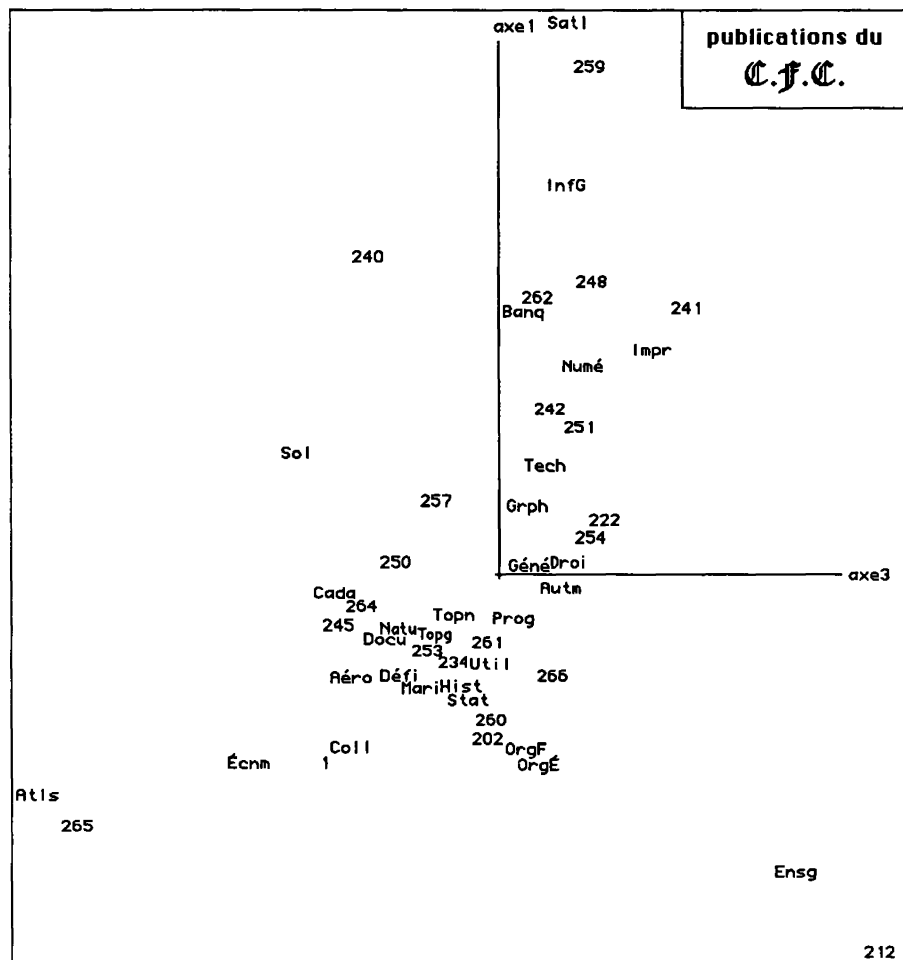
3 Analyse de la correspondance entre formes et thèmes

144 mots du lexique x 29 blocs thématiques

trace : 4.742e+0

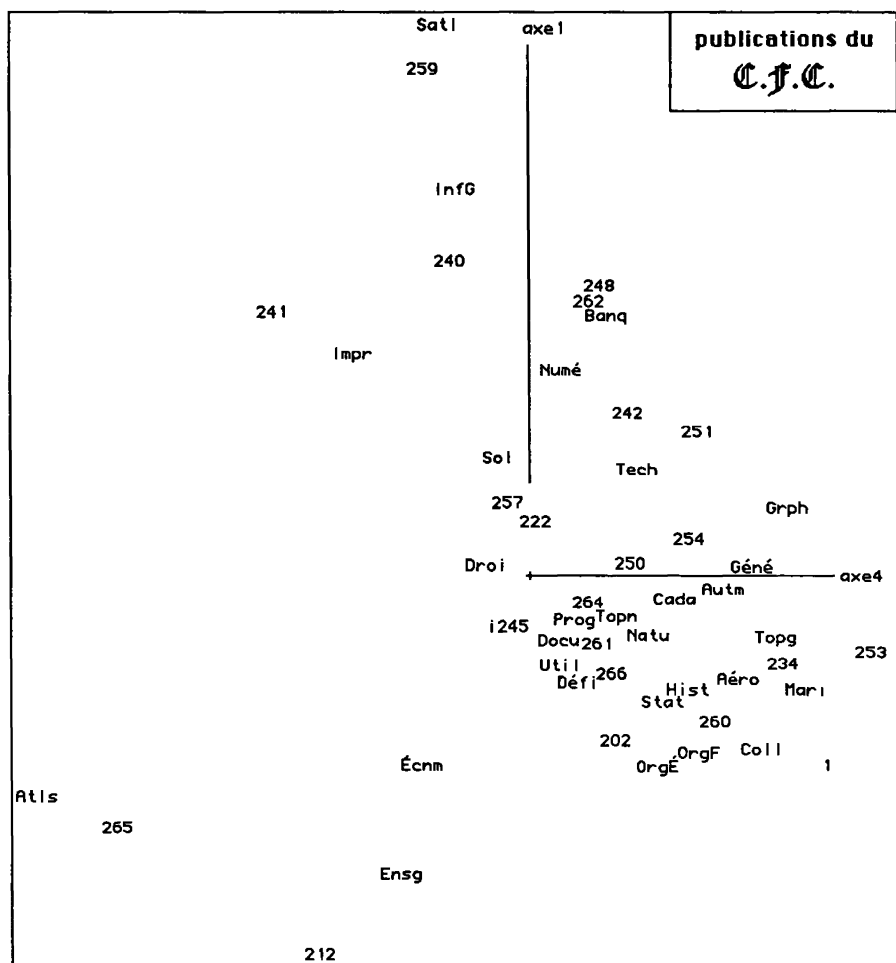
rang :	1	2	3	4	5	6	7	8	9	10	11
lambda:	4890	4667	4189	3171	2786	2442	2339	2047	1845	1778	1630 e-4
taux :	1031	984	883	669	587	515	493	432	389	375	344 e-4
cumul :	1031	2015	2899	3568	4155	4670	5163	5595	5984	6359	6702 e-4

Dans le plan (1, 2), on a un nuage allongé suivant l'axe 1, et dont se détache seulement, vers (F2>0), {Impr, 241}; suivi, de loin, par 251, {couleur, procédé,...} qui suit, comme il est logique, Impr.



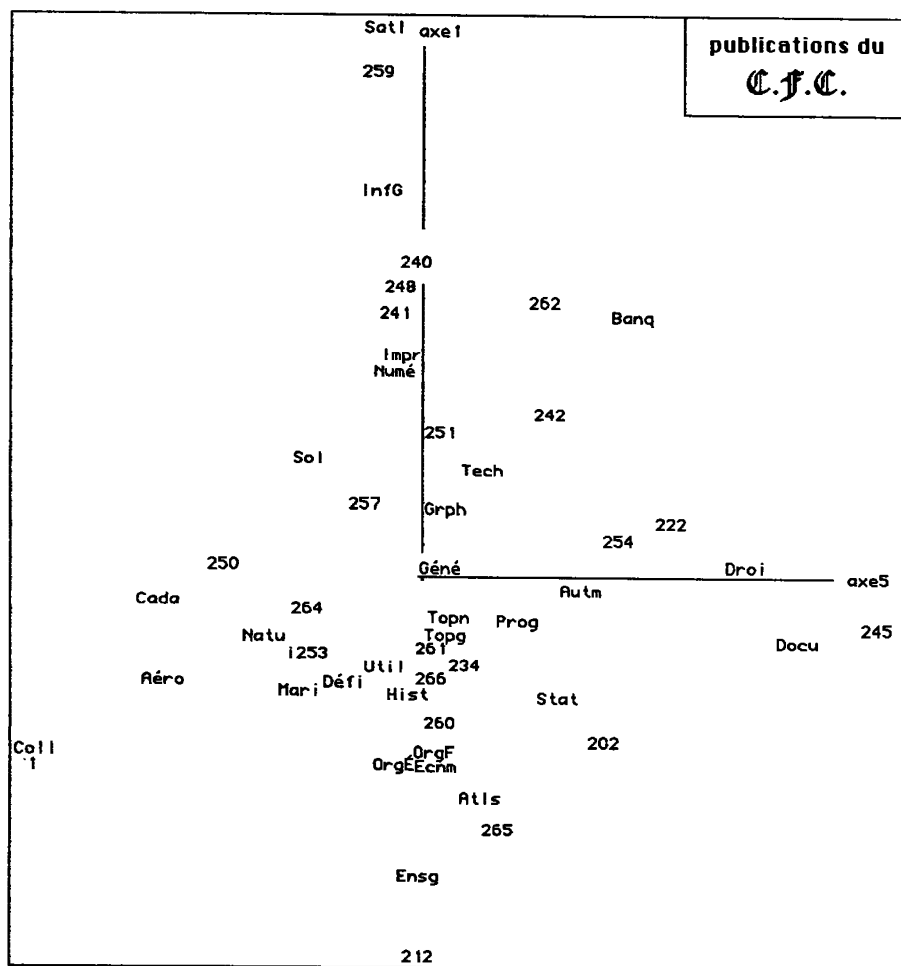
On présente les plans avec le nuage des 29 blocs thématiques, et les 22 centres de classes de mots, marqués par leurs numéros. L'axe 1 reçoit 74% de son inertie de l'opposition entre {Satl, InfG}, techniques nouvelles, ($F1 > 0$), d'une part, et Ensg, problème classique, ($F1 < 0$), d'autre part; ce qui permet d'interpréter $F1$ comme un facteur de technique.

Les deux thèmes Atls et Ensg, qui s'opposent sur l'axe 3, apportent à cet axe plus de 70% de son inertie. Seul Ensg a, sur le plan (1, 2), une qualité de représentation excellente: $COR1(Ensg) + COR3(Ensg) = 81\%$; pour Satl, on a: $COR1(Satl) + COR2(Satl) = 81\%$.



Opposés quant à F3, Atls et Ensg s'unissent pour apporter à l'axe 4, du côté ($F4 < 0$), 49% de son inertie. Dans le plan (3, 4), la qualité de représentation de Atls est de 68%.

Dans l'espace engendré par les axes 1 à 5, les thèmes dont la qualité de représentation dépasse 80% sont, rangés par ordre de qualité décroissante: {Impr, Ensg, Atls, Satl}; on a ensuite environ 50% pour {InfG, Docu}; puis 40% pour Coll, etc...



L'axe 5 reçoit 46% de son inertie de l'opposition entre Docu et Coll.

Alors que se multiplient les bases de données bibliographiques, nous croyons que la présente expérience contribue à préparer l'analyse de ces bases.

Source des données

Bulletin du Comité Français de Cartographie, 1958-1992, (132 fascicules).