

Un outil informatique pour l'analyse graphique de données

René Thoraval

▶ **To cite this version:**

René Thoraval. Un outil informatique pour l'analyse graphique de données. [Rapport de recherche] RT-0033, INRIA. 1984. inria-00071338

HAL Id: inria-00071338
https://inria.hal.science/inria-00071338

Submitted on 23 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

IRIA

CENTRE DE RENNES
IRISA

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(3) 954 90 20

Rapports Techniques

N° 33

UN OUTIL INFORMATIQUE POUR L'ANALYSE GRAPHIQUE DE DONNÉES

René THORAVAL

Avril 1984

IRISA

Campus Universitaire de Beaulieu
Avenue du Général Leclerc
35042 - RENNES CÉDEX
FRANCE
Tél. : (99) 36.20.00
Télex : UNIRISA 95 0473 F

Université de Nantes
Institut de Mathématiques et d'Informatique
2, Chemin de la Houssinière
44072 NANTES CEDEX

Publication Interne n°215

Rapport IMI-Info n° 14

UN OUTIL INFORMATIQUE POUR L'ANALYSE GRAPHIQUE DE DONNEES*

René THORAVAL

Juin 1983
60 pages

PRESENTATION

Ce rapport présente un logiciel d'aide au traitement graphique des données (ANAGRAPH), élément d'une chaîne globale d'aide au traitement.

Les traitements graphiques disponibles sont dûs à Jacques BERTIN. L'analogie entre ces traitements est mise en évidence. Des conséquences pédagogiques et informatiques en sont tirées.

Le logiciel peut, inséré dans un ensemble bureautique, permettre la création de graphiques corrects, lisibles et synthétiques (l'aspect "développement" est envisagé).

Le logiciel est particulièrement interactif ; le dialogue est basé avant tout sur l'image et organisé de façon pédagogique. Le but cherché étant d'aider l'utilisateur à traiter des données, à produire des graphiques, non d'agir aveuglément à sa place.

---

* Travail effectué dans le cadre du stage de DEA Informatique de l'Université de Rennes I à l'IMI de NANTES sous la direction de Michel LUCAS.

PAPIER RECUPERÉ ET RECYCLÉ

IN SHORT

This report presents a computer aided data analysis system (ANAGRAPH) as a part of a larger set.

The available processings are graphic ones and are due to Jacques BERTIN.

The analogy between these processings is placed in prominent position. Thus, pedagogical inferences as well as computer's ones are drawn from.

Inserted into an office's system, this software will enable anybody to produce sound, readable and synthetical business graphics. It is very interactive, conversation being based on image and pedagogically structured.

The system's purpose consists in lending its aid to the user but not in working instead of him.

TABLE DES MATIERES

INTRODUCTION

I. EXEMPLES D'ETUDES DEBOUCHANT SUR DES GRAPHIQUES

1. Exemple 1 : la maintenance informatique

- l'enquête
- la présentation des résulats
- « les fournisseurs donnant le plus satisfaction »
- l'état du parc : « équipements et contrats de maintenance »
- une autre approche

2. Exemple 2 : l'interpréteur APL de MULTICS testé sur différentes machines

- le test
- la présentation des résulats »REMARQUES
- une autre approche
- analyse multidimensionnelle « classique » et méthodes « Bertin » : quelques remarques

3. Exemple 3 : offres d'emploi dans l'informatique

- une construction graphique inadaptée
- une autre approche

4. Exemple 4 : la production de viande dans la C.E.E. en 1966

- « les diagrammes inutiles »
- une autre approche

5. Exemple 5 : évolution des admissions au bac dans l'académie de Nantes

- le document administratif
- une autre approche

II. NECESSITE D'UNE METHODOLOGIE GLOBALE DU TRAITEMENT DES DONNEES

1. Le traitement des données : un problème global

2. Présentation incorrecte des données statistiques : l'absence de méthodologie globale

3. Une méthodologie globale : celle de BERTIN

III. UN LOGICIEL GLOBAL D'AIDE AU TRAITEMENT DES DONNEES

1. Un logiciel global

2. Un logiciel d'aide au traitement

3. Un logiciel très interactif : l'utilisateur est l'acteur du traitement

4. Un dialogue basé sur l'image

5. De nouveaux traitements graphiques

6. Multiplicité des traitements/ stabilité es regroupements

7. La place du présent travail

IV. ANALYSE DES TRAITEMENTS GRAPHIQUES PROPOSES PAR BERTIN

1. Graphique statique/ graphique dynamique

2. La graphique dynamique

3. Les traitements matriciels

- la matrice ordonnable
- le fichier-image
- le fichier-matrice
- la matrice pondérée
- l'éventail de courbes

4. Remarque fondamentale : la matrice ordonnale st le traitement matriciel de base

V. LE CONCEPT DE TRAITEMENT MATRICIEL

1. Le concept de traitement matriciel

2. Exemples de traitement

- en matrice rdonnable
- en matrice pondérée
- en éventail de courbes

3. Schéma : analogie des traitements/ différences

4. Conséquences « pédagogiques »

5. Conséquences informatiques

- l'organisation du dialogue
  - un vaste « jeu » de procédures minuscules

VI. UN LOGICIEL STRUCTURE SELON LES CHOIX DE L'UTILISATEUR

1. Priorité à l'utilisateur/ priorité au dialogue

2. La hiérarchie fondamentale et son expression formelle

3. L'organisation des choix (niveau upérieur)

- analogie des traitements : paramétrage
- hiérarchiser les choix
- séparer organisation formelle des choix et réalisation effective

4. L'organisation de l'exécution des choix (niveau inférieur)

- une organisation hiérarchisée
- l'exécution numérique (niveau supérieur)
- l'exécution graphique (niveau inférieur)

VII. DESCRIPTION DU LOGICIEL ET COMMENTAIRES

1. Introduction

2. Formalisme adopté

3. Description

- G1/C1 : avant le traitement proprement dit
- G2/C2 : choix d'un traitement graphique
- G3/C3 : structure du traitement
- G4/C4 : phase statique du traitement
- G5/C5 : le niveau O de la phase dynamique
- G6/C6 : le niveau 1 de la phase dynamique
- G7/C7 : manipulations élémentaires

VIII. L'EXECUTION DES CHOIX DE L'UTILISATEUR

1. Introduction

2. L'exécution numérique

- phase statique
- phase dynamique

3. L'exécution graphique

- deux optiques
- le choix de l'optique « regénération de la partie à modifier »
- assimiler le traitement en matrice ordonnable à un traitement de texte
- matrice pondérée, éventail de courbes

IX. L'ASPECT «DEVELOPPEMENT »

1. Ce qu'implique un développement

- une phase d'expérimentation
- des corrections
- la possibilité d'implanter le logiciel sur des matériels variés
- offrir une gamme de versions plus ou moins évoluées

2. Ce qui peut le faciliter

3. Un matériel adapté
- au niveau graphique : un dispositif câblé de génération de caractères raphiques
- au niveau du dialogue

4. L'écriture du logiciel

X. LES PROLONGEMENTS A ENVISAGER

1. L'état des réalisations

2. Les prolongements
- choix des paliers
- reclassements automatiques
- traitements graphiques classiques
- vers l'amont : transformation du tableau de données, calcul des indicateurs
- vers l'aval : le « graphique de communication »

3. La communication des résultats

4. Les choix de codage visuel : définition es paliers
- le choix des paliers : un problème central de l'analyse des données
- ce que signifie le choix des paliers
- les problèmes posés
- la redéfinition dynamique des paliers : le propre des versions les plus évoluées du logiciel
- formaliser le choix des paliers : vers le calcul automatique de nouveaux indicateurs

XI. CADRE D'UTILISATION DU LOGICIEL

1. Une utilisation « grand public »

2. Insertion dans un ensemble bureautique pour la production de « business graphics »

3. Utilisation pédagogique
- pour la découverte de l'analyse de données
- pour la pédagogie d'autres domaines

4. Aide aux traitements statistiques

BIBLIOGRAPHIE

INTRODUCTION

Le recours aux données et à l'analyse de données est de plus en plus fréquent :
- traitement d'immenses tableaux statistiques à la manière de l'INSEE, certes,
- mais aussi, **prolifération** dans les journaux et revues, spécialisés ou non, dans les apports les plus divers produits par les entreprises, les administrations, etc., de tableaux de données, souvent assez petits et de graphiques (incluant ce qu'on appelle aux Etats-Unis les « **business graphics** », relatifs à la gestion d'une entreprise, à l'état d'un marché, etc).

Pourtant, le plus souvent, les **résultats** statistiques y sont **mal présentés** :
- démarche mathématique incorrecte,
- faible lisibilité,
- absence de point de vue synthétique.

Quelques exemples vont permettre d'illustrer ces remarques.

Ce sera aussi l'occasion d'introduire **une autre approche de la présentation des résultats**, approche basée sur les principes énoncés par Jacques BERTIN dans son ouvrage : «La graphique et le traitement graphique de l'information» [BER77].

I. EXEMPLES D'ETUDES DEBOUCHANT SUR DES GRAPHIQUES

1. Exemple 1 : la maintenance informatique

L'enquête [TR81a]

La revue « Temps Réel » a lancé en février-mars 1981 une enquête auprès de ses lecteurs sur la maintenance informatique. Elle en donne les principaux résultats dans son numéro du 22.06.81 [p.37 à 44].

Plus de deux cents lecteurs ont répondu : 166 questionnaires ont finalement été dépouillés et exploités.

L'enquête abordait les questions suivantes :
- répartition des établissements ayant participé à l'enquête,
- fonction des utilisateurs ayant complété les questionnaires,
- coût de la maintenance,
- comment les utilisateurs envisagent-ils l'évolution de ces coûts,
- les fournisseurs donnant le plus satisfaction en termes de maintenance,
- les équipements et les contrats de maintenance,

etc

La présentation des résultats

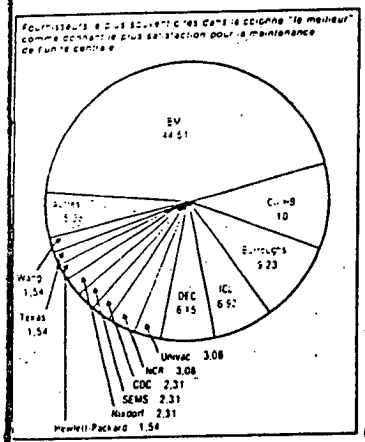
Ils sont présentés sur huit pages comprenant dix huit tableaux de chiffres, onze courbes, dix diagrammes de répartition, neuf diagrammes circulaires (« en fromage »), complétés par du texte.

Ils semblent extrêmement riches : la revue assure que « l'ensemble [des résultats globaux] constitue un outil unique d'évaluation des services de maintenance informatique ».

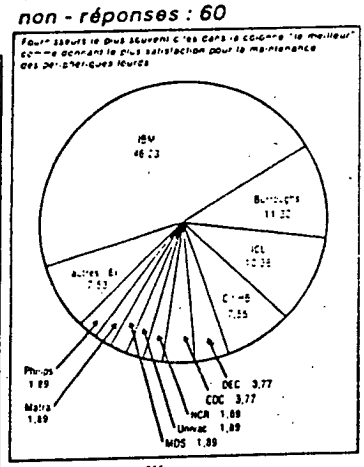
Pourtant, la lecture de l'article laisse le lecteur sur sa faim : **la richesse des résultats n'est pas mise en valeur par la présentation choisie.** Le lecteur regrettera que les dix huit tableaux numériques n'aient pas été remplacés par des constructions graphiques adaptées (quitte à ce qu'ils soient mentionnés en « notes », en caractères réduits)(voir par exemple Document 2). Il regrettera aussi que le sens de certains « caractères » ou « objets » n'ait pas été mieux défini : par exemple, qui sont les « autres fournisseurs » (voir Document 1,Document 2)?, qu'il n'ait pas été plus homogène d'un chapitre à l'autre : « logiciel d'application constructeur » + « logiciel d'application SSCI » (Document 1), « logiciel d'application » (Document 2), etc..

LA MAINTENANCE INFORMATIQUE

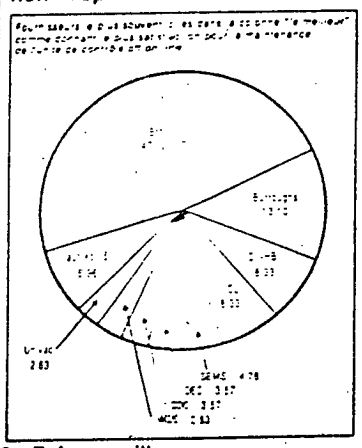
6 - 1. Les meilleurs en maintenance unités centrales
Citations : 130,
non - réponses : 36



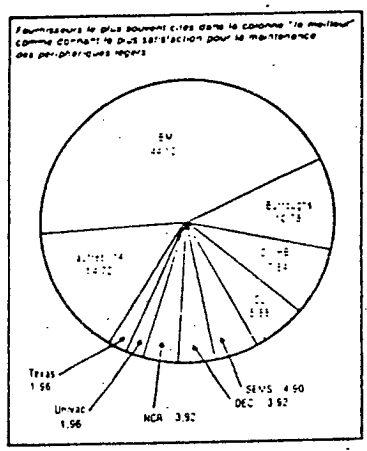
6 - 2. Les meilleurs en maintenance périphériques lourds
Citations : 106,
non - réponses : 60



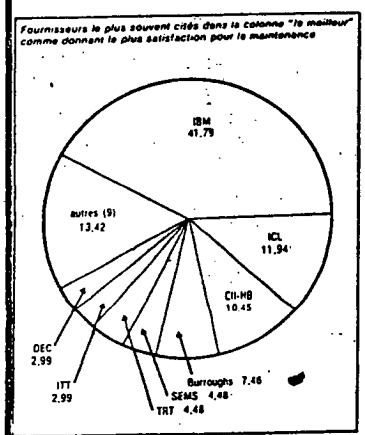
6 - 3. Les meilleurs en maintenance unité de contrôle off line / on line
Citations : 84,
non - réponses : 82



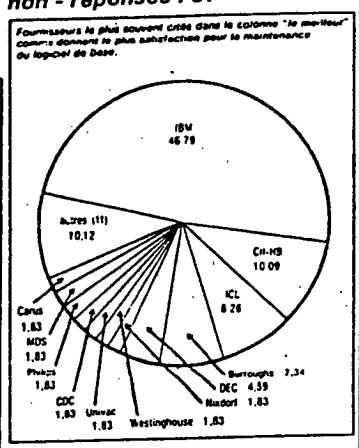
6 - 4. Les meilleurs en maintenance périphériques légers
Citations : 101,
non - réponses : 65



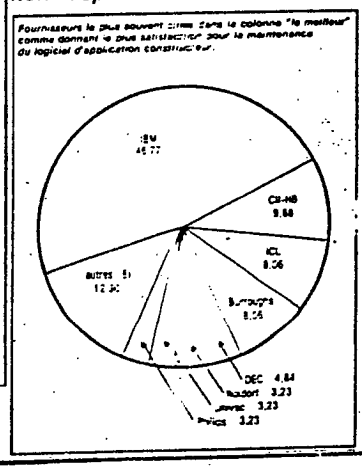
6 - 5. Les meilleurs en maintenance périphériques de transmission
Citations : 67,
non - réponses : 99



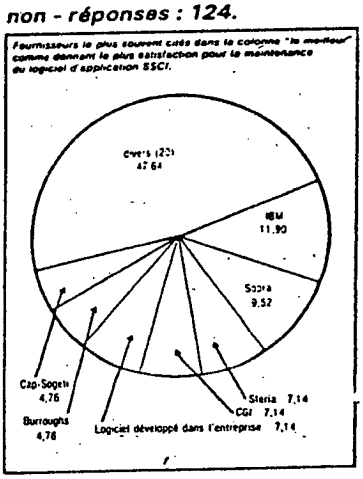
6 - 6. Les meilleurs en maintenance logiciel de base
Citations ; 109,
non - réponses : 57



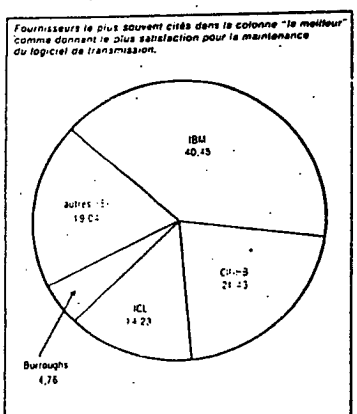
6 - 7. Les meilleurs en maintenance logiciel d'application constructeur
Citations : 62,
non - réponses : 104



6 - 8. Les meilleurs en maintenance logiciel d'application SSCI
Citations : 42,
non - réponses : 124.



6 - 9. Les meilleurs en maintenance logiciel de transmission
Citations : 42,
non - réponses :124



Les fournisseurs donnant le plus satisfaction : Burroughs et IBM

Conclusions

Pour la maintenance du matériel, IBM est le plus souvent cité dans la colonne « le meilleur », mais c'est bien sûr parce qu'il domine le parc. Et si l'on pondère en considérant les citations comme deuxième et comme troisième, **Burroughs** qui est le plus constamment apprécié. En logiciel, IBM est à nouveau en tête — vu son parc —, sauf dans les logiciels d'application où apparaissent les grandes SSCI, même si l'on peut encore noter à leur égard une certaine hésitation des utilisateurs. Et, on soupçonnera à nouveau la satisfaction des responsables — de même degré à l'égard des trois principaux fournisseurs — pour ce qui concerne la maintenance des logiciels de transmission. Enfin, CII-HB, le seul français qui pourrait faire pièce aux constructeurs étrangers, semble réellement avoir une mauvaise image — sauf pour la maintenance des logiciels de transmission. Constat hélas souvent entendu qui, sans doute, explique et entraîne les fréquentes hésitations des responsables informatiques à l'égard de CII-HB, par peur des problèmes et des risques.

Temps réel
22·6·81

DOCUMENT 1

Equipements et contrats

Pour préciser l'état du parc, et de la maintenance de ce parc, les questions 3, 4 et 5 visaient à connaître, par site, le nombre d'unités centrales de périphériques et de logiciels d'application ainsi que, dans chaque cas, le nombre et la nature des contrats de maintenance associés à chacun des éléments.

Tableau 7 : Equipements et contrats dans le secteur "industrie" Unités centrales

Constructeurs	nombre d'unités centrales	sur n... sites	nombre de contrats standard matériel	sur n... sites	nombre de contrats standard logiciel	sur n... sites	nombre de contrats particuliers	sur n... sites
IBM	61	41	61	31	44	29		
CII-HB	25	10	15	11	10	7		
Hewlett-Packard	16	3	6	4	3	3		
DEC	11	6	10	6	3	3	1	1
Burroughs	7	6	7	7	4	4		
Nixdorf	6	4	6	4	6	4		
ICL	5	4	5	4	1	1	1	1
MDS	4	1	4	1				
Wang	4	3	3	3	1	1		
CMC	3	1	3	1				
Olivetti	2	2	3	3	2	2		
Sems	2	1	2	2				
Data 100	2	1	1	1				
Itel	2	1	1	1				
Univac	2	2	2	2	2	2		
autres	20	11	10	10	8	8	11	3

Tableau 8 : Equipements et contrats dans l'industrie Périphériques

constructeurs périphériques lourds	nombre	sur n... sites	nombre de contrats standard	sur n... sites	nombre de contrats particuliers	sur n... sites
IBM	[illegible]	41	73	33	3	2
CII-HB	[illegible]	10	18	10	0	
Hewlett-Packard	[illegible]	3	3	3	0	
Burroughs	[illegible]	6	14	6	0	
DEC	[illegible]	6	7	4	2	2
BASF	[illegible]	2	4	2	0	
ICL	[illegible]	4	7	4	3	2
ITEL	[illegible]	1	1	1	0	
Nixdorf	[illegible]	4	5	4	0	
MDS	[illegible]	1	0		0	
Univac	[illegible]	2	2	2	0	
Memorex	[illegible]	3	6	3	0	
Texas	[illegible]	2	2	2	0	
Wang	[illegible]	2	2	2	0	
Sems	[illegible]	1	1	1	0	
autres	[illegible]	7	7	6	0	
périphériques légers						
IBM	[illegible]	34	111	34	0	
ITT	[illegible]	6	8	6	0	
CII-HB	[illegible]	7	18	7	1	1
MEMOREX	[illegible]	2	3	2	0	
ICL	[illegible]	4	6	4	1	1
Burroughs	[illegible]	6	25	6	0	
DEC	[illegible]	5	3	3	1	1
Univac	[illegible]	1	1	1	0	
Hewlett-Packard	[illegible]	2	2	2	0	
Data General	[illegible]	1	1	1		
MDS	[illegible]	1	0		0	
Nixdorf	[illegible]	3	2	2	1	1
autres	[illegible]	17	17	17	0	
périphériques de transmission						
IBM	[illegible]	15	29	15	0	
SAT	[illegible]	2	1	1	1	1
ICL	[illegible]	3	4	3	1	1
Burroughs	[illegible]	4	4	4	0	
CII-HB	[illegible]	2	11	2	0	
Olivetti	[illegible]	1	1	1	0	
Texas	[illegible]	1	1	1	0	
Nixdorf	[illegible]	1	1	1	0	
ITT	[illegible]	2	1	1		
SPI	[illegible]	1	1	1		
autres	[illegible]	6	6	6	0	

Tableau 9 : Equipements et contrats dans l'industrie Logiciels d'application

logiciels d'applications	nombre	sur n... sites	nombre de contrats standard	sur n... sites	nombre de contrats particuliers	sur n... sites
IBM	46	22	20	16	2	2
CII-HB	8	5	4	4	0	
Nixdorf	8	2	2	2		
Burroughs	6	3	1	1	0	
Sopra	6	6	3	3	0	
ICL	5	3	3	3	0	
CGI	4	3	2	2	1	1
Texas	4	2	0		0	
Steria	4	3	3	3	0	
SPI	4	2	2	2	0	
Data Pack	3	2	3	2	0	
Cap Sogeti	2	2	1	1	1	1
NCR	2	1	0		0	
autres	67	18	20	15	14	5

Abondance de contrats...

L'impression générale qui se dégage de ces chiffres est la profusion de contrats de toutes sortes pour tous les équipements (en particulier chez IBM), et le flou dans lequel semblent se trouver les utilisateurs ; à preuve ceux qui, citant un élément IBM, ne lui font pas correspondre un contrat de maintenance, alors qu'ils paient à coup sûr une redevance de maintenance...

Temps réel
22·6·81

Mais ce qui le frappera surtout, c'est **l'absence de présentation synthétique des résultats**. Il aura l'impression que l'enquête a été lancée, les réponses reçues et comptabilisées, puis publiées selon le découpage a priori du questionnaire. Pourtant, un tel découpage ne s'avère pas nécessairement pertinent a posteriori. En effet, la définition et le calcul de nouveaux indicateurs, combinant les anciens (i.e. les questions posées) auraient été plus éclairants, plus synthétiques.

Le lecteur constatera **qu'il n'y a pas eu de véritable mise en forme des résultats pour leur communication.**

Illustrons ces remarques sur deux exemples.

« Les fournisseurs donnant le plus satisfaction »

La question posée était : « Indiquez les trois fournisseurs qui vous donnent le plus satisfaction en termes de maintenance » [citer un « meilleur », un second, un troisième].

Les résultats sont présentés à travers neuf graphiques « en fromage » accompagnés, chacun, de commentaires particuliers (voir Document 1 ; les commentaires partiels n'y sont pas reproduits). Ils sont complétés par une introduction et une conclusion globales... déconnectées des remarques partielles ! A juste titre, elles constatent que « les réponses ne reflètent pas qu'un indice de satisfaction ; elles traduisent aussi l'état du parc » et qu'il faut donc les pondérer « en considérant les citations comme deuxième et comme troisième ».

Il convient de noter, en considérant cet exemple, **qu'il n'y a pas de présentation graphique synthétique. La synthèse est renvoyée au texte.** Pourtant un graphique bien construit serait bien plus parlant qu'un texte ! Cette remarque s'applique à la quasi-totalité des constructions graphiques proposées dans les publications.

L'état du parc : « équipements et contrats de maintenance »

La lecture de la conclusion de l'étude précédente amène à prendre en compte les résultats décrivant la répartition des « équipements et contrats de maintenance » entre les fournisseurs.

Malheureusement, **les données sont hétérogènes** : le parc considéré n'est plus que celui du secteur « industrie » alors que l'étude précédente se référait à l'ensemble : -industrie - commerce et distribution - services - administrations et collectivités locales- ; les types d'équipement sont définis différemment, tout comme les « autres » fournisseurs.

Ce qui frappe surtout, c'est le recours à une présentation des données numériques brutes, éclatées en cinq tableaux. La lisibilité des résultats est très mauvaise. **L'avalanche de chiffres n'autorise que des remarques très anecdotiques** (voir Document 2).

Aucun effort de communication n'a été réalisé.

Une autre approche

Le Document 3 résulte d'une autre approche de la présentation des résultats statistiques, approche inspirée des travaux de J. BERTIN.

Les données dispersées dans les neuf diagrammes « en fromage » (cf. Document 1) ont été réunies pour constituer un tableau « les meilleurs en maintenance ». Les données consignées dans les cinq tableaux du Document 2 ont été synthétisées dans le tableau « Equipements et contrats (standard) » (les « contrats particuliers » n'ont pas été pris en compte).

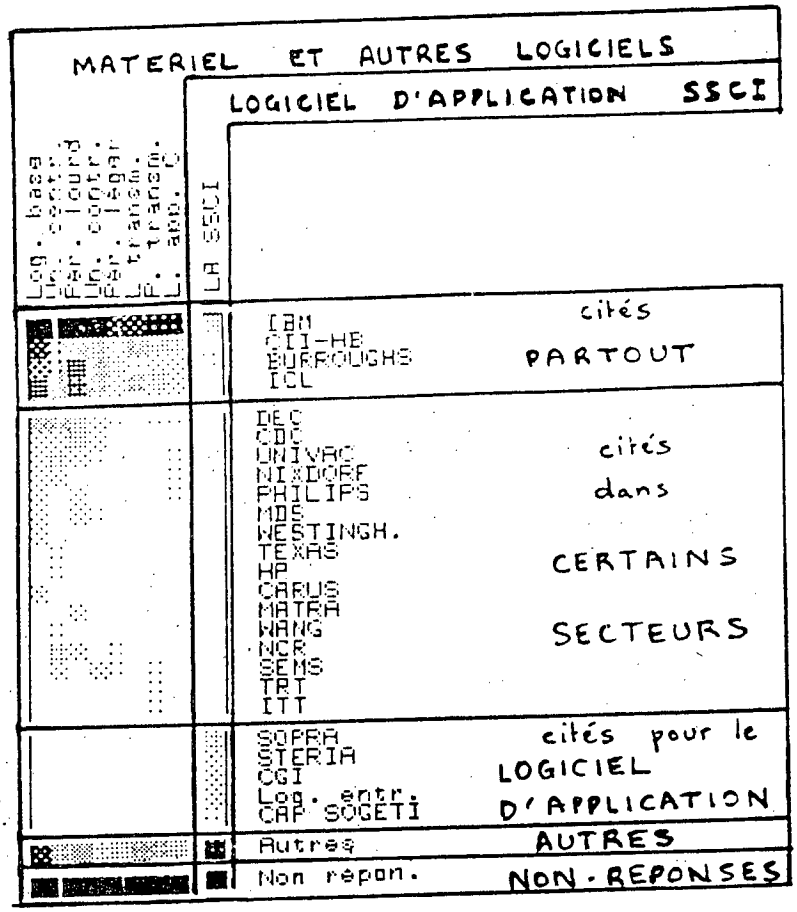
Ces deux tableaux de données numériques ont été traités à l'aide du logiciel comme « matrices ordonnables » (voir IV.3.,V.2.). Les résultats des traitements ont été mis en valeur sous forme de « graphiques de communication ».

Que pouvons-nous constater en voyant le Document 3? En admettant que :

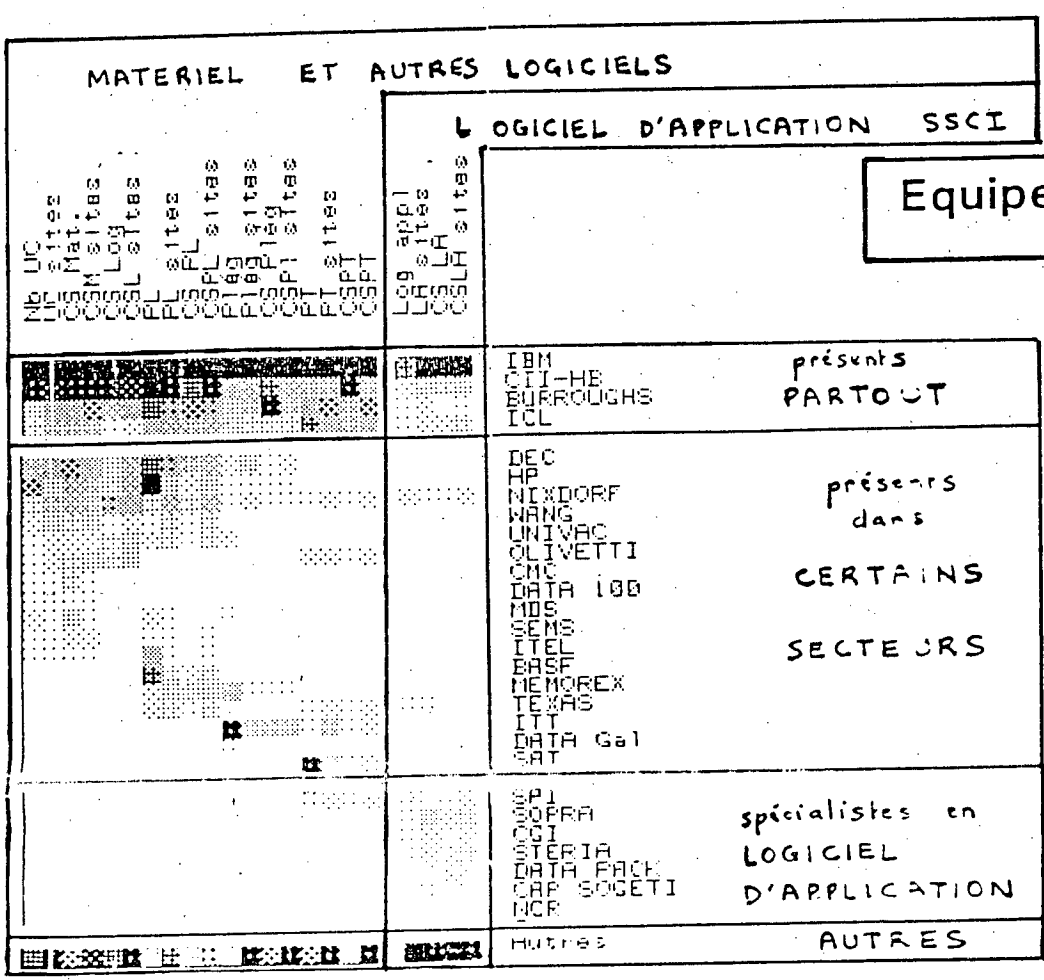
- l'état du parc dans le secteur « industrie » ne diffère pas fondamentalement de l'état du parc global,
- la non-homogénéité de la rubrique « autres » n'importe pas,

nous percevons d'emblée synthétiquement la corrélation forte entre « les meilleurs en maintenance » et les principaux fournisseurs.

Ainsi, la conclusion tirée par la revue est confirmée. Elle apparaît **visuellement** sans nécessiter de discours supplémentaire.



LES MEILLEURS EN MAINTENANCE
(nombre de citations)



Equipements et contrats
(standard)

Nous pourrons simplement regretter que « Temps Réel », au vu de sa conclusion, n'ait pas jugé bon de définir de nouveaux indicateurs permettant par exemple de voir en quoi « le meilleur en maintenance » différait parfois du principal fournisseur. Là aurait résidé l'**information utile** au lecteur. Un indicateur intéressant aurait pris en compte les « deuxièmes » et « troisièmes » « meilleurs » ainsi que le rapport : pourcentage des citations comme « le meilleur »/pourcentage d'occupation du parc.

2. Exemple 2 : l'interpréteur APL de Multics testé sur différentes machines

Le test [CANAL83]

Le numéro 95 de la revue du CIC de Grenoble contient un article de B. RAPPACHI et Y. SIRET. Ces derniers ont soumis l'interpréteur d'APL de MULTICS aux tests de la firme HARRIS, « test bien connu des spécialistes ». « Ces tests ont pour but de déterminer, ur différentes machines, les temps moyens d'exécution d'un certain nombre d'instructions supposées représentatives. L'intérêt de ce test est précisément de voir se rapprocher des machines très dissemblables en ...prix ou de différencier des machines ayant des vitesses élementaires très voisines ».

La présentation des résultats (Document 4)

Les auteurs nous offrent un vaste tableau numérique : - en lignes : les machines - en colonnes : les instructions - (voir Document 4).

Ils ajoutent : « pour tirer des conclusions synthétiques [!] de ce tableau déjà éloquent [!] nous avons utilisé le logiciel EDA et le programme BANACA de la bibliothèque ADDAD pour la représentation graphique ». (Voir Document 4).

Remarques

Il est clair que le tableau de données numériques, aussi « éloquent » soit-il ne peut fournir que des **informations anecdotiques**. L'anecdote peut être marquante : la machine la plus rapide effectue l'instruction MBSP en un temps cinquante-six fois moindre que la machine la plus lente! Cependant, elle n'est pas resituée dans une perspective globale...ni même résituable.

Quant à la représentation graphique (issue sans doute d'une analyse en composantes principales), elle comporte ces deux défauts graves :

- elle n'est pas suffisamment « parlante » du premier coup d'œil : il faut véritablement **« lire » le graphique** et éventuellement se référer au « mode d'emploi » que les auteurs jugent utile d'y ajouter ;
- ses **qualités de synthèse** sont **très limitées**; ces limites sont celles du « plan principal » ; ce qui se passe dans d'autres plans n'apparait plus graphiquement...mais dans le texte.

Une autre approche

Ici encore, le tableau des données numériques a été traité à l'aide du logiciel en « matrice ordonnable ».

Il suffit de regarder (Document 5):

- pour voir l'opposition entre les machines très rapides (IBM 370/168, AMDHAL 470V/6) et les machines très lentes (HB68, DECSYSTEM 2020 et 2040);
- pour remarquer l'opposition entre la grande masse des instructions et deux instructions « bizarres » :MBSP (« Membership ») et TRSP (« Transposition »).

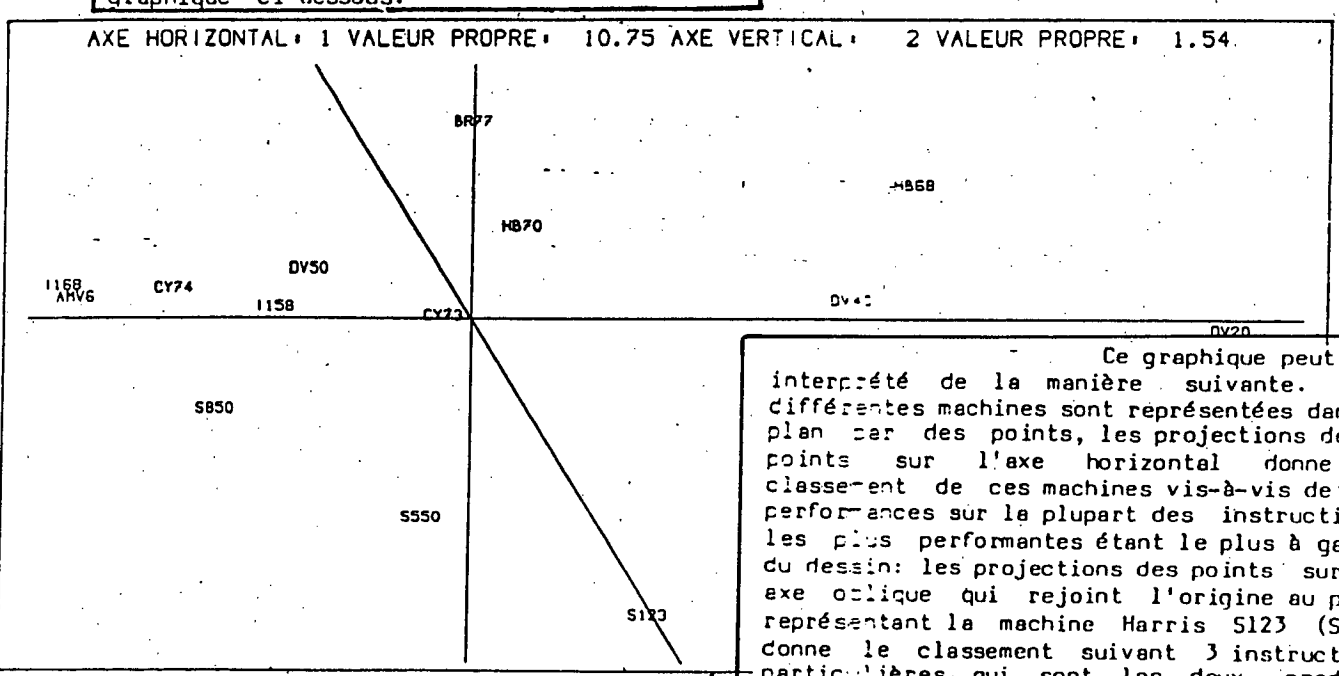
L'anecdote évoquée plus haut prend maintenant son sens : l'instruction MBSP (comme TRSP) a un comportement distinct de la masse des autres : dispersion bien plus grande des vitesses d'exécution des différentes machines.

Le graphique proposé sur le Document 5 aurait pu remplacer avantageusement : - **et** les données numériques - **et** l'essentiel des cent cinquante lignes de texte - ...**et** la représentation graphique produite par BANACA. Les données numériques auraient pu figurer cependant en « notes », en petit, comme « matériau » disponible aux lecteurs qui voudraient en savoir plus par eux-mêmes...et donc poursuivre le traitement.

L'INTERPRETEUR APL DE MULTICS TESTE SUR DIFFERENTES MACHINES

	PLRD	LGRD	MXRD	EXPN	VABS	INDX	SORT	TAKE	MOSP	TRSP	OPCH	OPIN	INRN	MTDV	FIBO
S123	1.9	0.7	2.3	57.2	3.6	5.3	52.9	1.0	38.9	138.5	21.1	53.9	7.0	10.0	349.4
S550	1.3	0.4	1.6	39.8	2.2	3.6	34.8	0.6	19.4	96.7	14.7	37.7	4.7	6.9	264.0
S850	0.6	0.2	0.8	25.4	1.5	2.1	19.5	0.4	17.2	46.9	6.4	23.6	2.3	4.0	159.8
I158	0.9	0.5	1.1	42.7	3.5	1.1	14.7	0.9	61.5	4.2	5.7	22.3	4.1	5.2	196.7
I168	0.3	0.2	0.5	9.0	1.5	0.4	4.7	0.4	23.5	1.3	2.0	6.9	1.0	1.5	79.6
HB70	1.5	1.1	2.3	40.0	2.9	4.4	28.1	1.4	625.5	27.9	9.3	29.0	5.9	23.0	365.2
HB68	2.5	1.9	4.0	61.0	5.1	7.7	49.7	2.5	970.7	50.0	15.5	43.6	10.1	37.2	650.0
DV20	3.5	6.8	5.3	103.9	8.1	5.4	56.0	2.8	504.3	82.1	35.8	54.6	12.3	33.6	1534.6
DV40	2.5	5.1	4.6	43.1	6.2	4.3	56.0	2.2	441.2	37.4	18.6	47.8	7.7	11.4	1162.7
DV50	1.2	2.3	1.9	18.7	3.1	1.6	14.8	1.0	111.0	14.8	5.8	16.9	3.2	5.3	486.1
CY73	2.0	0.5	1.7	40.4	1.8	4.5	40.4	0.9	263.6	11.3	7.0	37.1	6.6	6.4	319.0
CY74	0.9	0.3	0.8	9.0	0.8	1.9	21.0	0.5	142.3	4.6	3.0	16.8	3.2	2.4	155.8
BR77	4.6	0.8	1.7	25.4	1.6	3.2	32.8	1.3	17.4	24.8	3.2	13.2	9.9	31.7	271.4
AMV6	1.0	0.1	0.6	4.0	0.4	0.4	4.9	0.1	80.7	0.7	7.8	12.1	1.4	1.0	55.0

Pour tirer des conclusions synthétiques de ce tableau déjà éloquent nous avons utilisé le logiciel EDA et le programme BANACA de la bibliothèque ADDAD pour la représentation graphique ci-dessous.

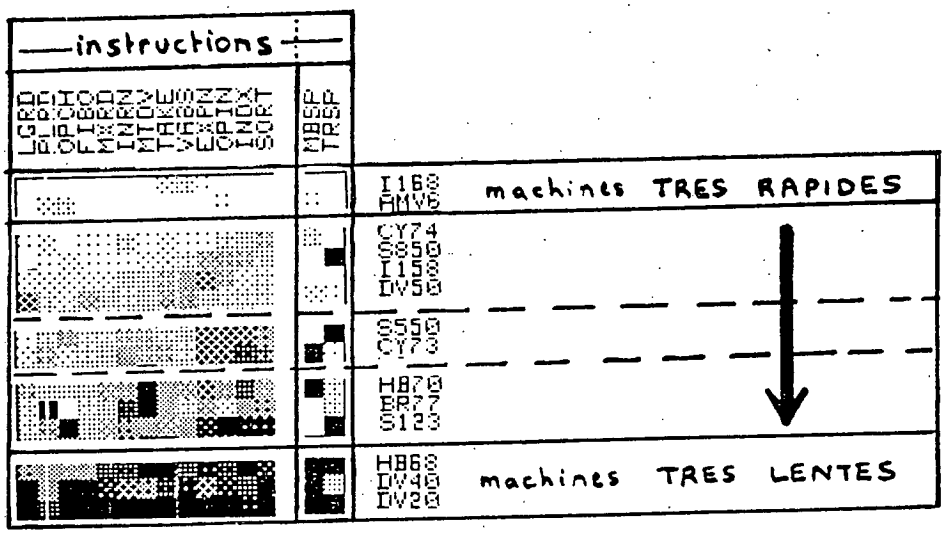


Notes : La matrice MR est singulière. Nous l'avons rendue invisible en forçant ses valeurs diagonales à 1000.0, ce qui ne modifie pas l'esprit du test qui est de chercher un temps moyen pour la "division matricielle".
En comparant les temps des 2 types de CPUs de CII-HB, on retrouve le facteur voisin de 1.63, qui est le rapport actuel de "puissance" entre les 2 machines.

Ce graphique peut être interprété de la manière suivante. Les différentes machines sont représentées dans un plan par des points, les projections de ces points sur l'axe horizontal donne le classement de ces machines vis-à-vis de leur performances sur la plupart des instructions: les plus performantes étant le plus à gauche du dessin: les projections des points sur un axe oblique qui rejoint l'origine au point représentant la machine Harris S123 (S123) donne le classement suivant 3 instructions particulières qui sont les deux produits extérieurs (caractères et entiers) et la transposition, les machines les moins bien classées étant justement du côté de la Harris S123 (qui est maintenant obsolète). Des études plus détaillées effectuées grâce à EDA mais sur lesquelles nous ne attarderons pas ici montrent que la réalité est un peu plus compliquée, ces machines se séparent en 4 groupes.

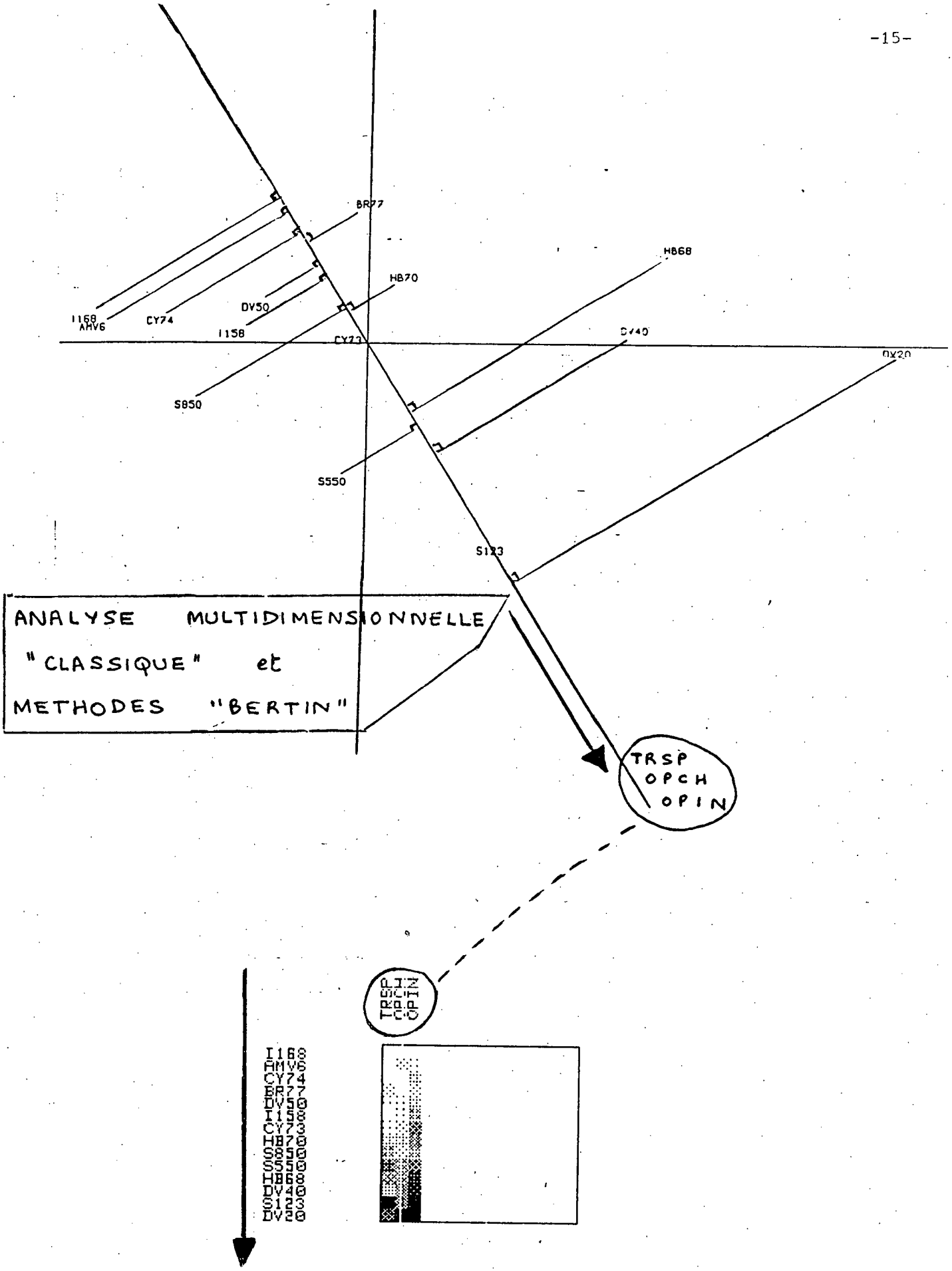
Le CANAL (CIC Grenoble)
mai 83 n° 95

L'INTERPRETEUR APL DE MULTICS SUR DIFFERENTES MACHINES



le CANAL
(CIC Grenoble)
Mai 83
N° 95

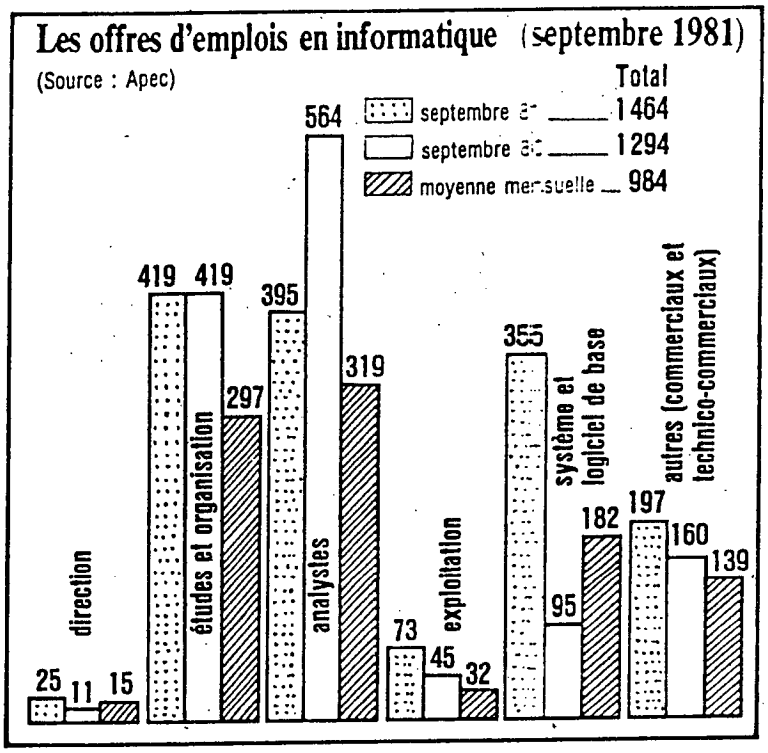
DOCUMENT 5



DOCUMENT 5 bis

12 ter

EVOLUTION DES OFFRES D'EMPLOI DANS L'INFORMATIQUE

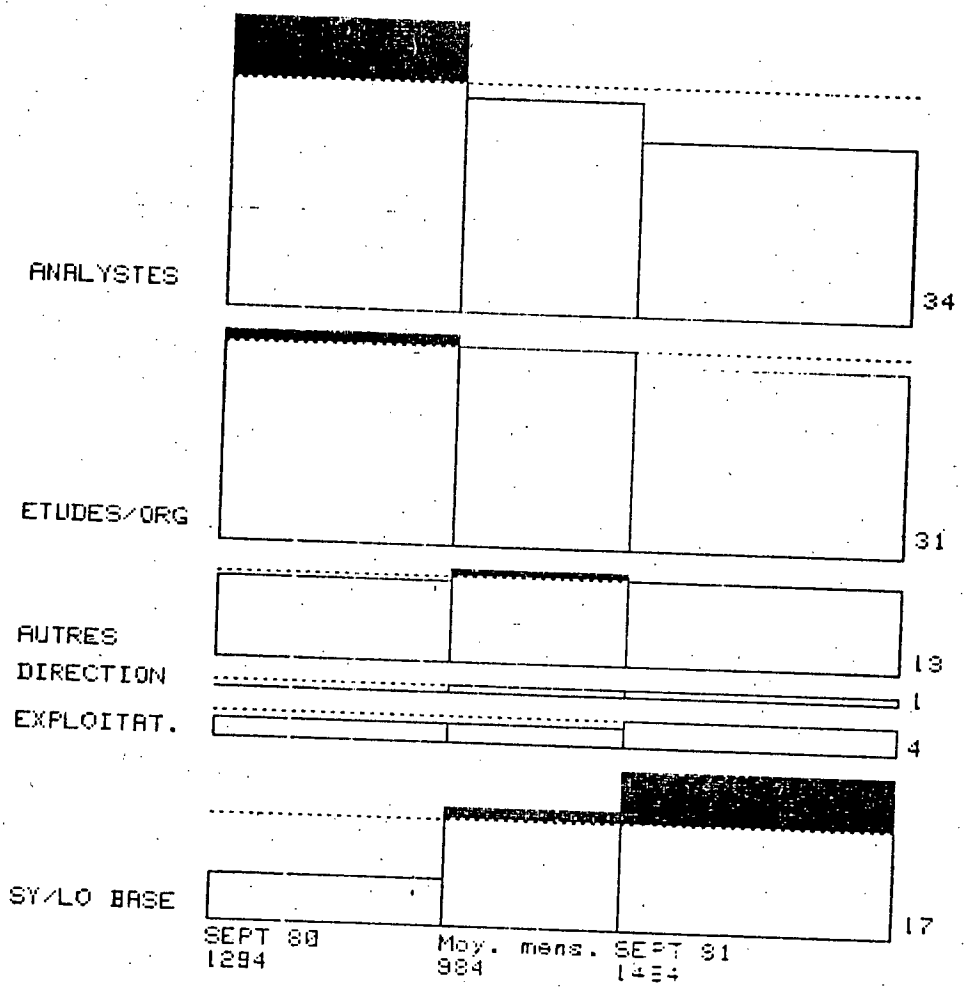


Temps réel

2.11.81

DOCUMENT 6

EVOLUTION DES OFFRES D'EMPLOI DANS L'INFORMATIQUE



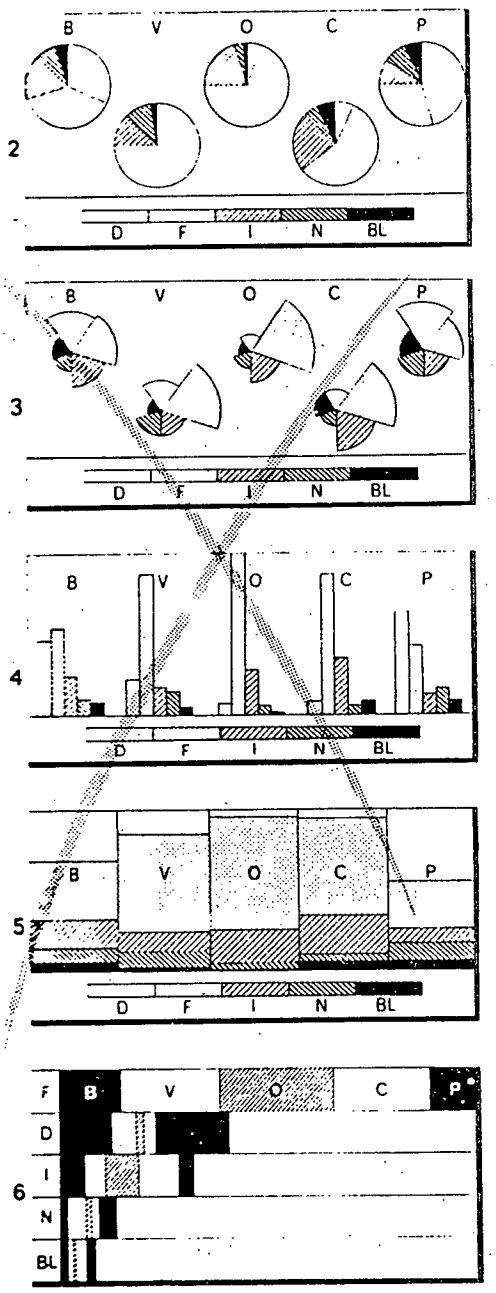
Les offres d'emploi en informatique
TEMPS REEL (2.11.81)

LA PRODUCTION DE VIANDE DANS LA C.E.E. (1966)

1

	B	V	O	C	P
D	32	15	5	6	45
F	38	60	70	60	29
I	17	12	20	24	9
N	7	10	4	4	11
BL	6	3	1	6	6
	100	100	100	100	100

données numériques



extrait de [BER 77]

Analyse multidimensionnelle « classique » et méthodes « Bertin » : quelques remarques

- Le graphique du Document 4 peut selon les auteurs être interprété ainsi : « les différentes machines sont représentées dans un plan par des points, les projections de ces points sur l'axe horizontal donnent le classement de ces machines vis à vis de leur performance sur la plupart des instructions : les plus performantes étant le plus à gauche du dessin ». Le classement des machines tel qu'il apparaît sur le Document 5, issu d'un traitement « Bertin », est pratiquement identique.
- Les auteurs indiquent aussi que «les projections des points sur un axe oblique qui rejoint l'origine au point représentant la Harris S123 (S123) donne le classement suivant trois instructions particulières» ; il s'agit de : TRSP, OPIN et OPCH.

Un traitement « Bertin » ne prenant en compte que ces instructions, moyennant « masquage » des autres colonnes (cf. VII.C6 et Document 23) et une simple permutation des lignes, permet d'aboutir (voir Document 5 bis) à un classement très voisin.

- Ainsi, des résultats très comparables à ceux obtenus par des méthodes multidimensionnelles « classiques » peuvent l'être par des méthodes « Bertin ». Pour une amorce de la discussion des rapports entre ces méthodes : voir [BER77] p.251 à 253.

3. Exemple 3 : offres d'emploi dans l'informatique

Une construction graphique inadaptée [TR81b]

Dans son numéro du 2.11.81, la revue « Temps Réel » analyse le marché de l'emploi des informaticiens.

Pour décrire son évolution, elle propose le graphique du Document 6.

Malheureusement, **un tel graphique est « à lire » :** nous ne pouvons « voir » **d'emblée** l'information synthétique : **comparer les différentes professions, les différentes périodes**

Une autre approche

Les données numériques inscrites sur le graphique du Document 6 ont été réunies dans un tableau unique. A l'aide du logiciel, un traitement en « matrice pondérée » a été effectué (voir IV.3.,V.2.) (cas limite de possibilité de ce traitement).

D'emblée, nous pouvons « **voir** » (Document 7):

- l'opposition de l'évolution du marché des emplois d'analystes et de celui des professionnels en système et logiciel de base ;
- le volume des offres d'emploi dans chaque profession ;
- l'importance relative des offres d'emploi en septembre par rapport à la moyenne mensuelle.

4. Exemple 4 : la production de viande dans la C.E.E. en 1966

Cet exemple est repris de [BER77] : pages 194 et 195.

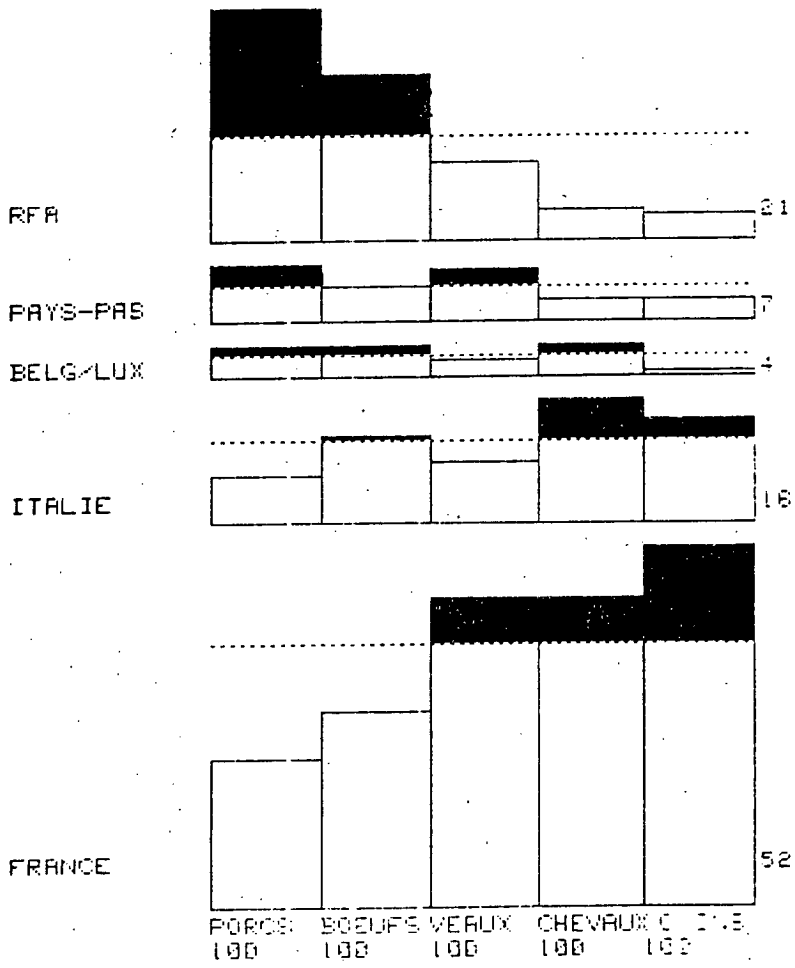
« Les diagrammes inutiles »

Le Document 8 présente une série de graphiques inutiles pour représenter l'information contenue dans le tableau numérique de gauche [pays : D = Allemagne, F = France, I = Italie, N = Pays-Bas, BL = Belgique + Luxembourg] - Viandes » B = boeufs, V = veaux, O = ovins, C = chevaux, P = porcs].

Inutiles ? « Les ministres qui étudient le marché de la viande sont-ils bien avancés quand ils ont sous les yeux la figure (2)? Ils n'en tirent que quelques rapports et le tableau (1) est beaucoup plus efficace. Comment se regroupent ou s'opposent les cinq nations dans le cadre de cette information sur la production ? Aucune des constructions (2) à (6) n'apporte la réponse ».

DOCUMENT 9

LA PRODUCTION DE VIANDE DANS LA C.E.E. (1966)



Production de viande dans la CEE (1966)

repris dans [BER 77]

EVOLUTION DES ADMISSIONS AU BAC

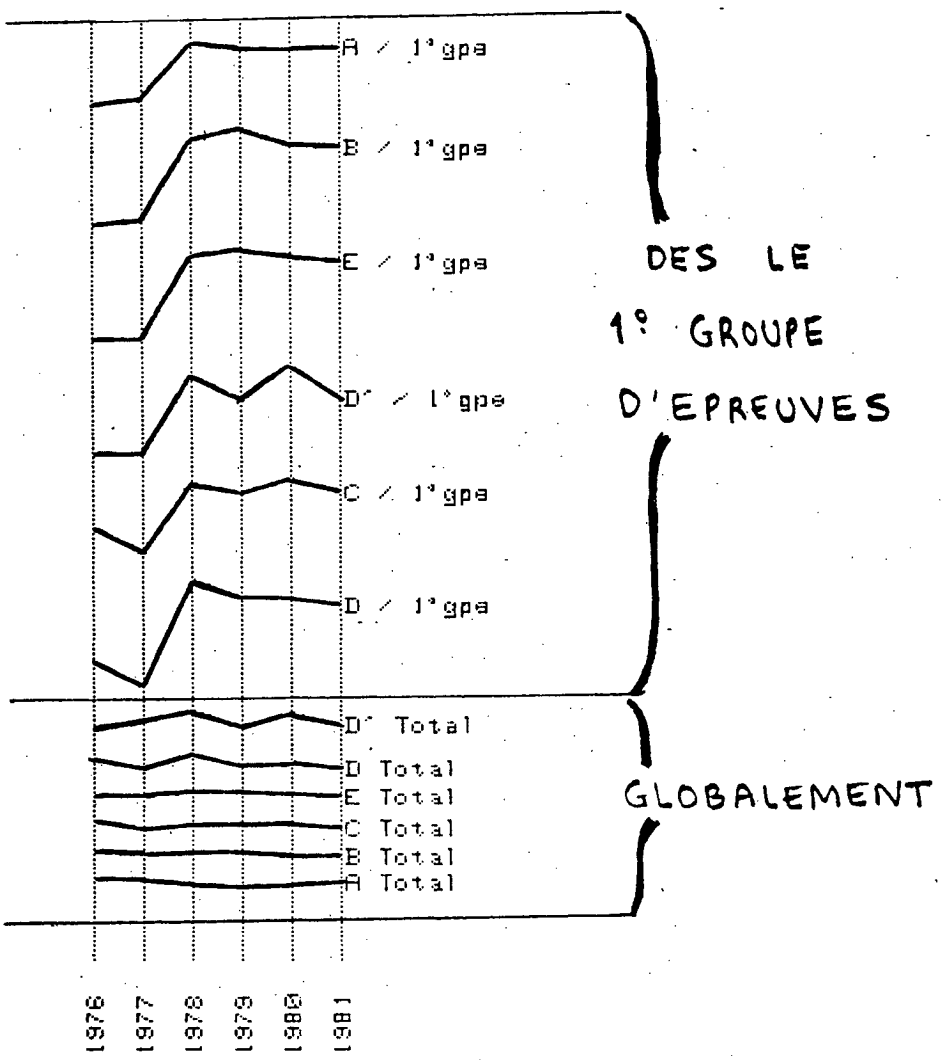
		1er groupe	1er & 2e groupe
		admis en %	TOTAL DES ADMIS en %
A	1981	39.4	65.4
	1980	38.8	63.6
	1979	38	61
	1978	42	65
	1977	19	69
	1976	17	70
B	1981	32.4	61.2
	1980	32.6	62.1
	1979	42	66
	1978	37	56
	1977	11	64
	1976	10	66
C	1981	38	65.6
	1980	46.2	71.3
	1979	38	68
	1978	43	70
	1977	15	67
	1976	23	74
D	1981	32.7	54.8
	1980	36.9	59.7
	1979	38	58
	1978	48	69
	1977	10	55
	1976	15	63
D'	1981	30.9	55
	1980	48.8	66
	1979	29	53
	1978	43	68
	1977	13	57
	1976	13	51
E	1981	38.3	61.7
	1980	40.4	64.4
	1979	45	67
	1978	42	66
	1977	12	63
	1976	12	61

Source : Rectorat de l'Académie de NANTES

RESULTATS COMPARES DU BACCALAUREAT

1976 - 1977 - 1978 - 1979 - 1980 - 1981

EVOLUTION DES ADMISSIONS AU BAC



ADMIS AU BACCALAUREAT

(Acad. de Nantes)

DOCUMENT 11

Une autre approche (Document 9):

Le traitement en « matrice pondérée » « fait apparaître l'information d'ensemble et révèle un problème proprement politique : les nations forment deux groupes, l'Allemagne et les Pays-Bas d'une part, l'Italie et la France d'autre part, qui ne peuvent avoir que des politiques opposées...ou complémentaires. Et ici l'Union belgo-luxembourgeoise joue le rôle d'arbitre ».

Le Document 9 a été obtenu à l'aide du logiciel.

5. Exemple 5 : évolution des admissions au Bac dans l'Académie de Nantes

Le document administratif [ACAD81]

Il présente sur deux pages l'évolution des admissions au Bac dans l'Académie de Nantes entre 1976 et 1981.

De ces deux pages, un sous-tableau a été extrait. Il est présenté sur le Document 10.

Il témoigne à la fois de la **richesse** des données brutes ccumulées par les services du rectorat et de l'**illisibilité** de leur communication.

Une autre approche

De ce sous-tableau, qui est en fait la réunion de six tableaux distincts, a été tiré un tableau unique.

A l'aide du logiciel, un traitement en « éventail de courbes » a été effectué (voir IV.3, V.2.).

Le Document 11 présente les résultats.

D'emblée, nous pouvons **voir** l'opposition très marquée entre l'évolution du taux d'admission dès le premier groupe d'épreuves et l'évolution du taux d'admission global, indépendamment des sections (A,B,...).

II. NECESSITE D'UNE METHODOLOGIE GLOBALE DU TRAITEMENT DES DONNEES

1. Le traitement des données : un problème global

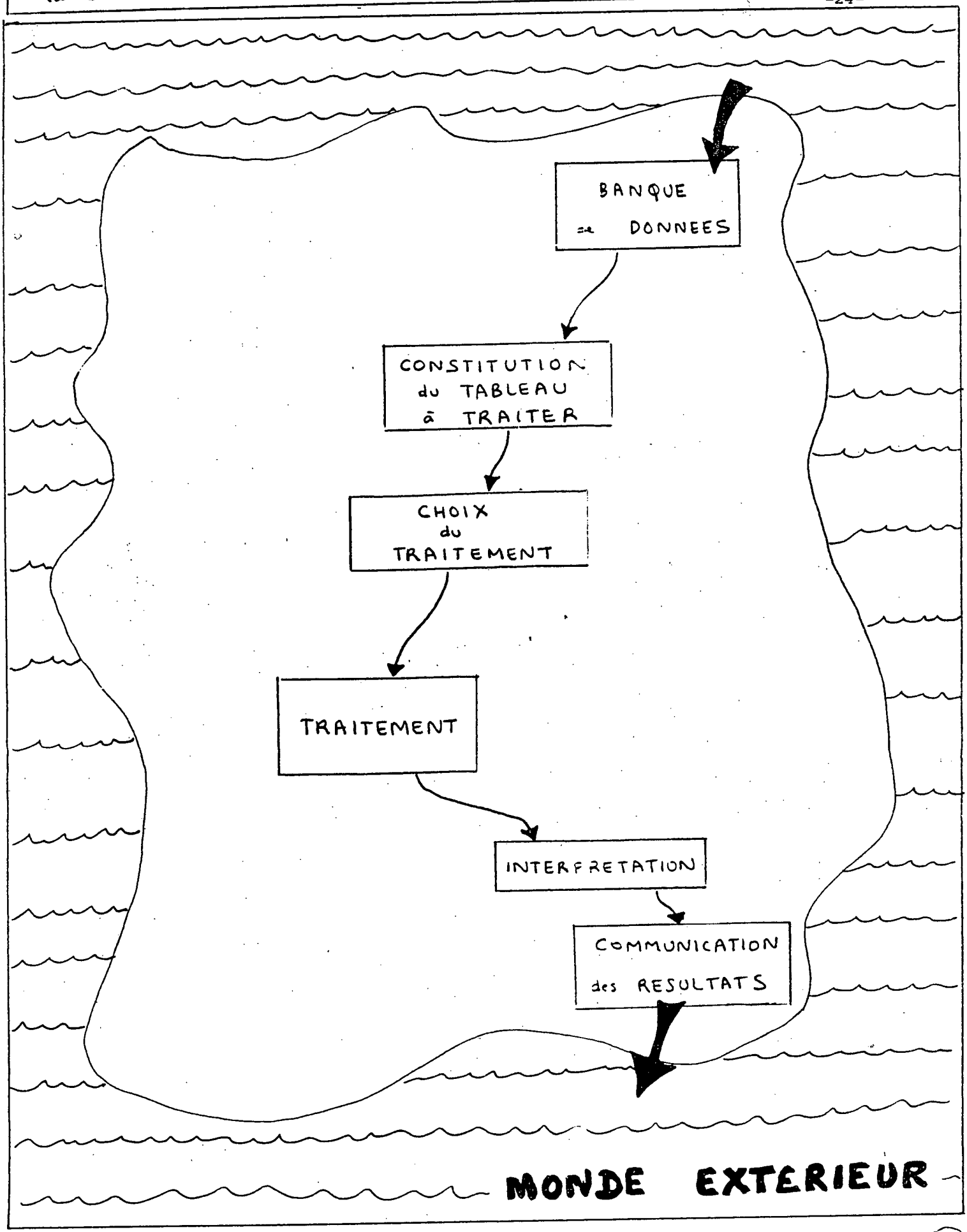
Le traitement des données est un problème global qu'on peut schématiser par la **chaîne de traitement** décrite sur le Document 12.

Cette chaîne va du « **monde extérieur** » dont on extrait les **données « brutes »**, inutilisables telles quelles, jusqu'au « monde extérieur » à nouveau, où elle apporte **l'information utile à la décision**, information qui ne prend son sens que dans et par ce « monde extérieur ».

Cette chaîne inclut une phase d'« analyse », de traitement proprement dit. Cependant, la réduire à cette seule phase parce qu'on dispose d'algorithmes et de calculateurs puissants et qu'on croit que cela suffit, expose à des dangers sérieux : calculs énormes sur des données dénuées de sens, traitement inadapté aux buts poursuivis, impossibilité de communiquer clairement les résultats, etc...

Chaque maillon de la chaîne a son importance, la procédure d'analyse n'est que l'un d'eux.

la "chaîne" de traitement des données



2. Présentation incorrecte des résultats statistiques : l'absence de méthodologie globale

A travers quelques exemples, nous venons de voir que les résultats statistiques présentés dans la plupart des publications le sont incorrectement.

Cela tient avant tout à l'absence d'une méthodologie globale soutenant ces présentations. Par « méthodologie globale », il faut entendre une méthodologie prenant en compte la globalité du problème du traitement des données, d'un bout à l'autre de la chaîne. Il s'agit notamment de ne pas négliger la phase de « communication des résultats » qui a une existence propre ; trop souvent il n'en est pas tenu compte ; d'où la présentation de tableaux de données numériques brutes, de graphiques « de travail » inadaptés à la diffusion des résultats, etc...

3. Une méthodologie globale : celle de Bertin

Jacques BERTIN, dans son livre cité précédemment ([BER77]), propose une méthodologie couvrant la globalité du problème du traitement des données. Le lecteur est invité à s'y reporter ; le présent travail s'appuie sur les principes proposés par BERTIN. Il vise à informatiser ce que lui et son équipe ont réalisé manuellement.

III. UN LOGICIEL GLOBAL D'AIDE AU TRAITEMENT DES DONNEES

Les propositions de Bertin, et tout particulièrement les traitements graphiques qu'il a imaginés, semblaient attendre le développement de l'**infographie interactive** : voir sa critique des systèmes informatiques existants et ses souhaits ([BER77] p.21). D'où l'idée de **réaliser un logiciel global** reprenant la méthodologie de Bertin, basé sur l'infographie interactive, et destiné à aider le traitement des données. C'est vers la réalisation d'un tel logiciel que s'oriente une partie des activités de l'Institut de Mathématiques et d'Informatique de Nantes.

1. Un logiciel global (Document 13)

Il couvre l'ensemble de la chaîne présentée sur le Document 12.

2. Un logiciel d'aide au traitement

Il s'agit d'**aider l'utilisateur** à traiter un ensemble de données, non de les traiter à sa place. L'utilisateur est toujours **l'acteur** des opérations, le système n'est là que pour l'assister, le décharger des âches laborieuses, exécuter ses ordres. Un tel logiciel n'est manifestement pas adapté à un utilisateur qui souhaiterait déposer ses données « brutes », presser un bouton et attendre les bras croisés, l'information utile à la décision ; un tel utilisateur, d'ailleurs, n'aurait rien compris au problème du traitement des données.

L'aide se situe à tous les niveaux (cf.Document 13) ; citons :

- aide à la constitution du tableau à traiter : tel indicateur a-t-il un sens mathématique ? Si oui, le calculer automatiquement ? un « point de comparaison » a-t-il été défini ? etc ;
- aide au choix du traitement quels sont les moyens de calcul disponibles ? Quel phénomène l'utilisateur veut-il analyser ? etc.;
- aide au traitement : ce sera tout particulièrement vrai pour les traitements graphiques originaux proposés par Bertin ainsi que nous le verrons ; c'est envisageable pour les traitements multidimensionnels classiques : par exemple, l'utilisateur pourra se prononcer pour ou contre la poursuite d'une analyse en composantes principales suivant le taux d'inertie déjà expliqué, les résultats pressentis, etc;

- aide à l'interprétation : elle résidera dans la structure même des traitements graphiques « Bertin » moyennant de faciliter la tâche de l'utilisateur (rédaction d'un document de travail, stockage/reprise d'un état graphique quelconque) ; pour les traitements multidimensionnels classiques, elle incluera la possibilité de générer automatiquement des graphiques du type de celui du Document 4 ; dans tous les cas, elle impliquera la présentation claire des hypothèses du traitement;
- aide à la communication : pour les résultats issus des traitements « Bertin », ce sera une aide à la mise en page (marges, tailles variées des caractères), au « filtrage » des résultats, à la cartographie (voir X.3.) ; dans les autres cas, il est possible de reprendre les résultats selon les traitements et présentations « Bertin », entre autres.

3. Un logiciel très interactif : l'utilisateur est l'acteur du traitement

Un lien très serré est établi à tous les niveaux de la chaîne entre celle-ci et le « monde extérieur », cadre et sens du traitement. Cette présence permanente de l'« information externe » au traitement, cette possibilité donnée à l'utilisateur d'orienter e traitement dans le moindre détail tout en le déchargeant des tâches laborieuses, lui permettent d'obtenir les résultats les plus fins i.e. les mieux adaptés à sa demande, les plus utiles à sa décision. (Voir Document 13).

4. Un dialogue basé sur l'image

« **Il faut réapprendre à VOIR** » nous dit Bertin. **VOIR**, car l'image permet souvent une communication plus simple, plus synthétique (cf.[BER77] p. 176 à 234 : « Sémiologie de la graphique »).

La communication entre l'utilisateur et le système sera onc aussi visuelle que possible : - nature même des traitements graphiques - recours à des schémas plutôt qu'à du texte (ex : « schéma d'homogénéité ») - référence directe à l'écran pour désigner une partie d'un graphique sans recours à un nom, à un numéro (voir VII.C7) - etc -.

5. De nouveaux traitements « graphiques »

Bertin a imaginé de nouveaux traitements graphiques permettant de **traiter synthétiquement et visuellement des tableaux de plus de trois caractères** auxquels se heurtait la « graphique » classique, « statique » (histogrammes, etc.). Il ne nie pas les « VERTUS » de cette « graphique » classique, il les remet simplement à leur place :voir [BER77] p.24 à 31 : « le synoptique des constructions graphiques ».

Il ne nie pas davantage les « vertus » des traitements multidimensionnels « classiques » (ACP, AFC, etc) mais il les analyse et les resitue (Voir [BER77] p.251 à 253).

Le logiciel contiendra donc : - les traitements graphiques : « statiques » (trois caractères au plus) et « dynamiques » (plus de trois caractères) - les traitements multidimensionnels classiques - ou du moins l'aide au choix de l'un de ces traitements. En effet, on peut concevoir, si l'on veut « développer » un tel logiciel sur micro-ordinateurs, que certains traitements multidimensionnels classiques, plus « lourds », soient confiés à l'« extérieur ».

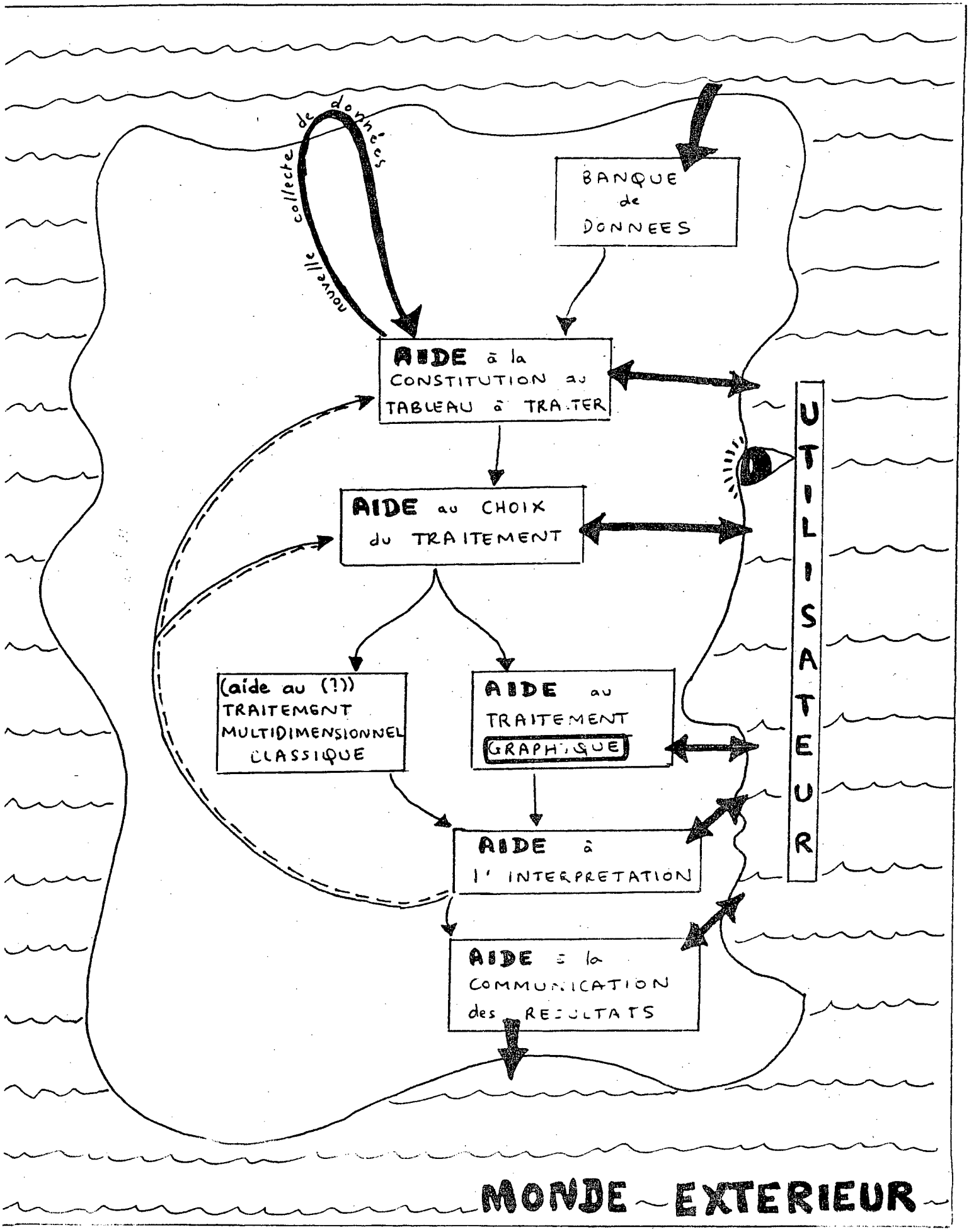
6. Multiplicité des traitements/ stabilité des regroupements

En analyse de données, aujourd'hui, on compte de moins en moins sur la découverte et/ou le choix d'un traitement « génial » unique...mais de plus en plus sur l'exécution d'une série de traitements sur les mêmes données et l'étude de la stabilité des regroupements ffectués d'un traitement à l'autre.

C'est pourquoi le logiciel permettra (et aidera) l'utilisateur, après interprétation des résultats d'un traitement, à **remonter « en arrière » dans la chaîne** :

- redéfinition du tableau de données par collecte de données nouvelles, choix de nouveaux indicateurs, etc;
- choix d'un autre traitement.

L'**aide** portera par exemple sur la définition de nouveaux indicateurs en réinvestissant les résultats concernant les choix de « paliers » dans le traitement précédent (Voir X.4.).


nouvelle collecte de données
BANQUE de DONNEES
AIDE à la CONSTITUTION au TABLEAU à TRAITER
AIDE au CHOIX du TRAITEMENT
(aide au (?)) TRAITEMENT MULTIDIMENSIONNEL CLASSIQUE
AIDE au TRAITEMENT GRAPHIQUE
AIDE à l' INTERPRETATION
AIDE à la COMMUNICATION des RESULTATS
UTILISATEUR
MONDE EXTERIEUR

7. La place du présent travail

Il se situe dans le cadre de la conception de ce logiciel global, essentiellement au niveau « **aide au traitement graphique** ».

Cependant, compte tenu de la nature même de ces traitements, **il touche à l'ensemble de la chaîne** :

- le module d'aide u choix du traitement est partiellement défini ; la partie « traitement graphique » l'est presque entièrement;
- la connexion entre « aide à la constitution du tableau de données » et le corps du traitement est esquissée grâce à la mise en évidence du lien entre « choix visuels » (dont choix des « paliers ») et définition/calcul des indicateurs;
- l'aide à l'interprétation est pour l'essentiel réalisée dans le cas des traitements graphiques (document de travail, stockage d'un état graphique);
- les principes du module d'aide à la communication des résultats obtenus par traitement graphique sont partiellement définis.

De plus, un certain nombre d'**idées,** de **remarques,** relatives à cette partie du logiciel, sont **valables pour la conception d'ensemble**, notamment celles sur la priorité donnée au dialogue, la structuration de ce dialogue et sa description formelle par une grammaire (Voir VI.1et2).

Enfin, pour obtenir des résultats pratiques, il a fallu réaliser des versions simplifiées de certains « maillons », ne serait-ce que celui, essentiel, de l'acquisition des données.

IV. ANALYSE DES TRAITEMENTS GRAPHIQUES PROPOSES PAR BERTIN

Je renvoie le lecteur, pour l'essentiel, au chapitre B : « Les constructions graphiques » de [BER77] et tout particulièrement au « Synoptique des constructions graphiques » (p.24 à 31).

1. Graphique statique/ graphique dynamique

Bertin différencie :

- la graphique « **statique**» adaptée aux tableaux d'au plus trois caractères ;
- la graphique « **dynamique** » indispensable quand on dépasse trois caractères. Un traitement dynamique se décompose en **deux phases** :
  - phase « statique » = génération u graphique initial;
  - phase « dynamique » = suite de reclassements à partir du graphique initial.

2. La graphique dynamique

Au sein de la graphique « dynamique », il oppose :

- les traitements « **matriciels** » où le graphique a un aspect « matriciel » calqué sur le tableau (matriciel) des données;
- les traitements « **non matriciels** » en « collections » (tableaux, cartes).

3. Les traitements matriciels

Quant aux traitements matriciels, il en distingue cinq :

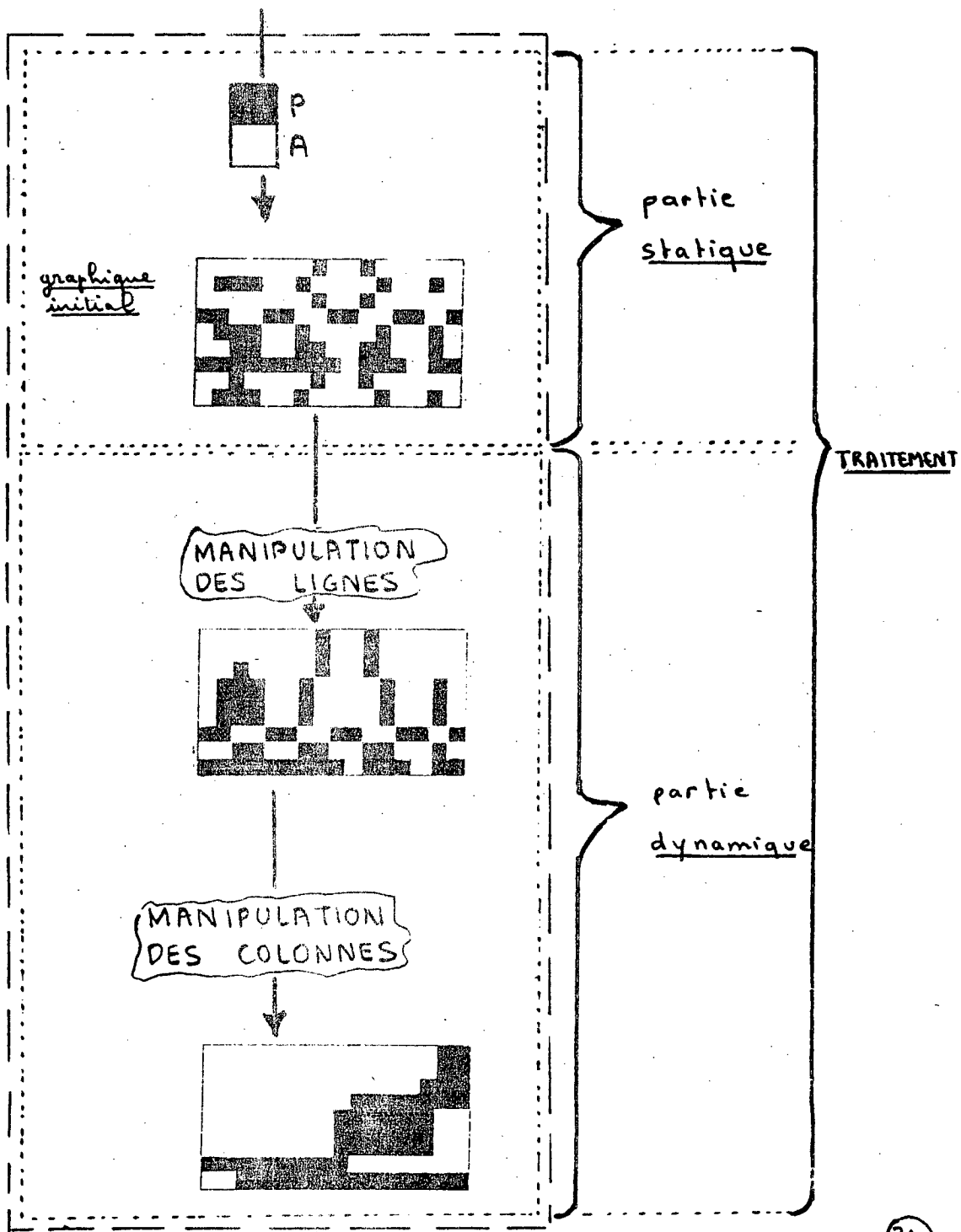
- la « **matrice ordonnable** » (Voir IV.3., V.2.);
- le « **fichier-image** » ne différant du précédent que parce que l'une des « composantes » (ligne ou colonne) est ordonnée a priori : seule l'autre sera donc manipulable;
- le « **fichier-matrice** s'applique à un tableau relevant normalement d'une matrice ordonnable mais dont l'une des dimensions est très grande : traitement d'un grand nombre d'objets à travers un petit nombre de caractères ; les manipulations (i.e. la « partie dynamique » du traitement) classent le fichier selon une succession de caractères;

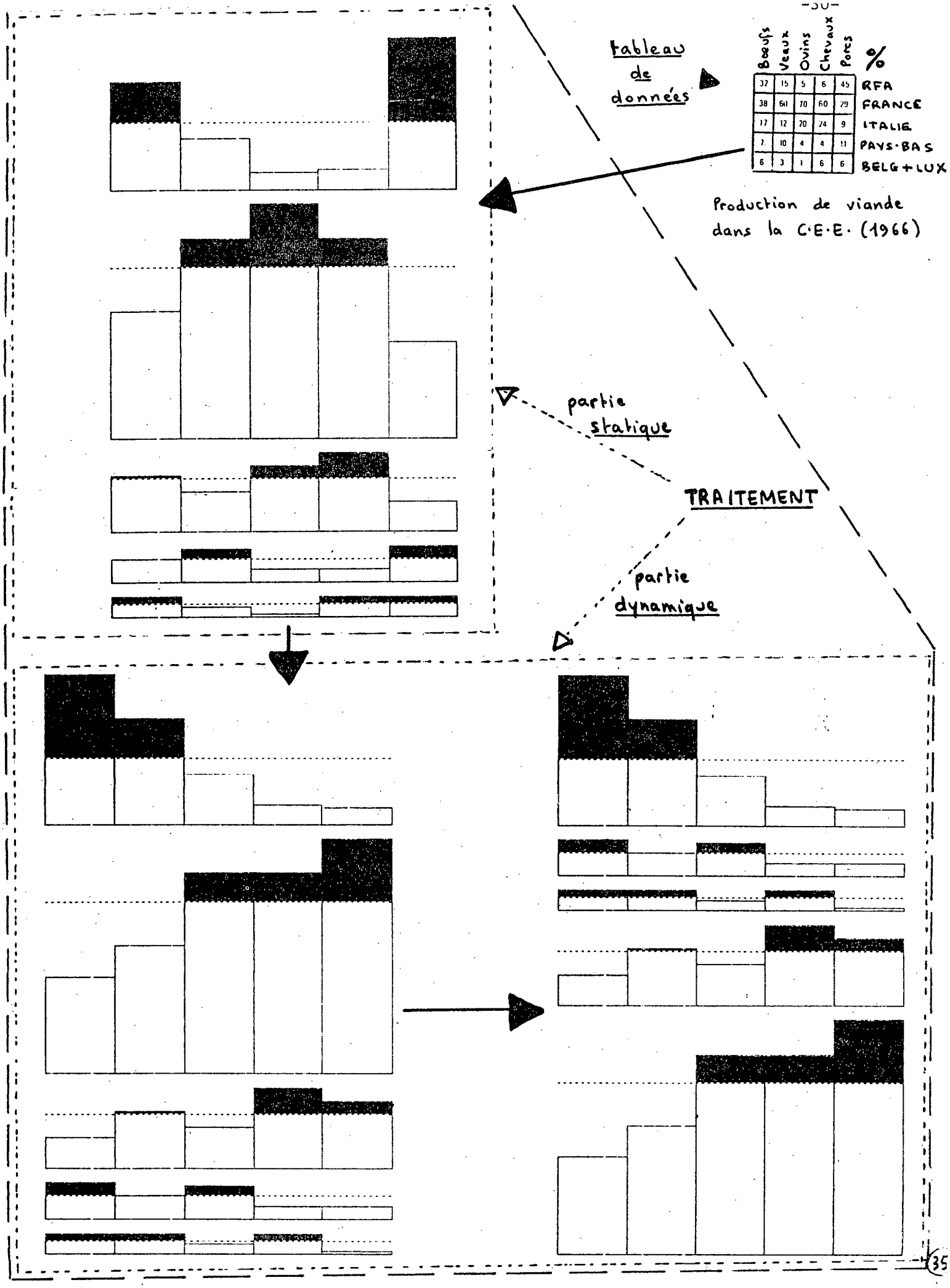
LA MATRICE ORDONNABLE

	⌀	U	∩	W	⊔	⊃	H	⊓	X	⌡	Σ	Z	O	⍺	⌀	⍺
COLLEGE	A	A	A	A	A	A	A	P	A	A	P	A	A	A	A	A
COOP. AGRIC.	A	P	P	P	A	A	P	A	A	A	A	P	A	A	P	A
GARE	A	A	A	A	A	A	A	P	A	A	P	A	A	A	A	A
EC. CL. UNIQ.	P	P	A	A	P	P	A	A	P	P	A	A	P	P	A	P
VETERIN.	A	P	P	P	A	A	P	A	A	A	A	P	A	A	P	A
MEDECIN	A	A	P	P	A	A	P	P	A	A	P	P	A	A	P	A
ADDUCT. EAU	P	P	P	P	P	P	P	P	P	A	P	P	P	A	P	P
GENDARM.	A	A	P	A	A	A	A	P	A	A	P	A	A	A	A	A
REMEMBR.	A	P	P	P	A	A	P	A	A	A	A	P	A	A	P	A

Services et équipements
(A = absent P = présent)

◀ Le tableau de données




tableau de données
Bœufs
Veaux
Ovins
Chevaux
Porcs
%
32 15 5 6 45 RFA
38 60 70 60 29 FRANCE
17 12 20 24 9 ITALIE
7 10 4 4 11 PAYS-BAS
6 3 1 6 6 BELG+LUX
Production de viande dans la C·E·E· (1966)
partie statique
TRAITEMENT
partie dynamique

- la « **matrice pondérée** » (voir IV.3., V.2.) ; mais le traitement évoqué sur le Document 15 ne peut s'appliquer valablement que si le tableau est de dimensions réduites ; sinon il sera traité en matrice ordonnable, les paliers étant définis en fonction des écarts à la moyenne car la matrice pondérée exprime, case par case, la différence entre répartition observée et répartition proportionnelle;
- l'**éventail de courbes** » (Voir IV.3., V.2.) ; pour pouvoir appliquer ce traitement, l'une des composantes doit être ordonnée a priori. Supposons qu'il s'agisse des colonnes : on considère alors des fonctions $z^k_j = f_k(x_j)$. L'éventail de courbes permettra de comparer leurs pentes : c'est son avantage sur le fichier-image. Mais pour y recourir, il faut que ces pentes aient une signification...

4. Remarque fondamentale : la matrice ordonnable est le traitement matriciel de base

Cela signifie que tout tableau de données qu'on peut traiter selon un traitement matriciel quelconque peut toujours l'être en matrice ordonnable. La réciproque est fausse.

Si le traitement en matrice ordonnable n'est pas toujours le plus adapté (ex : l'éventail de courbes est préférable pour comparer des pentes), il donne toujours cependant des résultats intéressants.

Par conséquent, une **version simplifiée du logiciel** ne comportant que ce traitement, est tout à fait concevable et viable : elle sera seulement moins souple, moins riche qu'une version complète.

V. LE CONCEPT DE TRAITEMENT MATRICIEL

1. Le concept de traitement matriciel

Tout traitement matriciel comporte deux parties :

- la **partie « statique »** débouchant sur la génération du **graphique initial** ; celui-ci a une forme « matricielle » car il est **calqué** sur le tableau (matrice) des données, les valeurs étant retranscrites « par une variation visuelle ordonnée du blanc au noir ou du plus petit au plus grand » ;
- la **partie « dynamique »** visant à faire apparaître les groupements d'objets, de caractères (le but de l'analyse de données) **en supprimant les distances visuelles inutiles** i.e. en rapprochant les lignes et/ou les colonnes qui se ressemblent visuellement. Traiter les données, c'est alors « simplifier sans détruire », grâce à la perception visuelle et spatiale.

2. Exemples de traitements

1. Le Document 14 présente schématiquement un traitement en **matrice ordonnable**.

Le tableau de données présente un ensemble de communes et les services et équipements disponibles dans chacune d'elles.

Le traitement est décomposé en ses deux parties. Ici, la règle de « **décalque** » du graphique initial est très simple car les données sont binaires : la séparation entre manipulation des lignes et des colonnes est schématique [Partie **statique** + partie **dynamique** : voir Document 14].

Une fois le graphique final obtenu, on peut faire apparaître les noms des communes et des services pour se livrer à l'interprétation.

2. Le Document 15 présente schématiquement un traitement en **matrice pondérée**.

C'est la reprise de l'exemple 4 (voir I.4.), les résultats sont communiqués dans le Document 9.

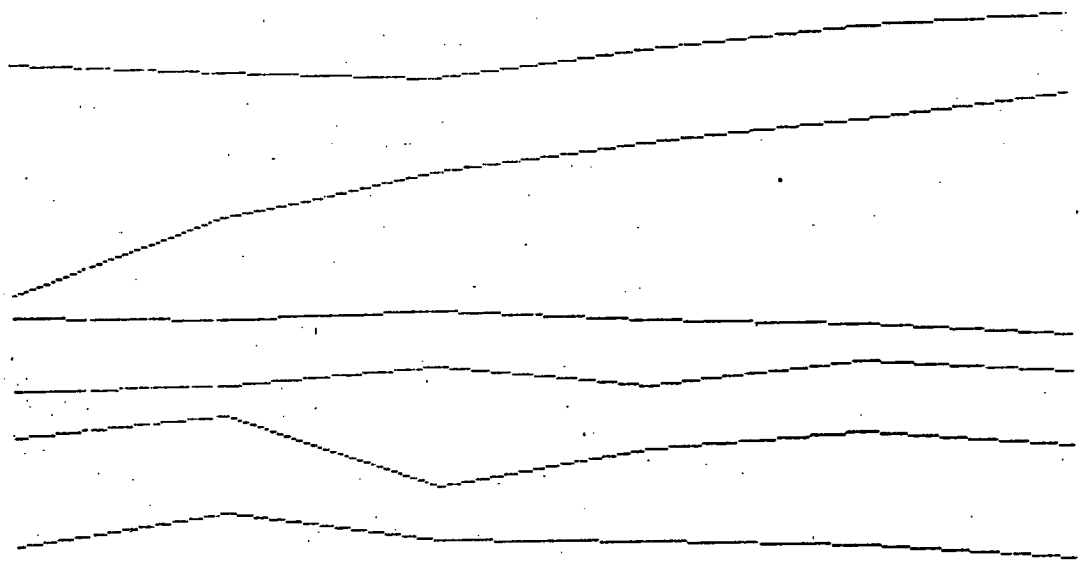
	76	77	78	79	80	81
A	3135	3005	2901	3170	3412	3526
B	1207	1559	1797	1959	2111	2269
C	1937	1906	1930	1855	1817	1727
D	4000	4053	4273	3937	4270	4064
D'	207	221	171	193	203	191
E	434	481	435	427	418	394

◀ tableau de données

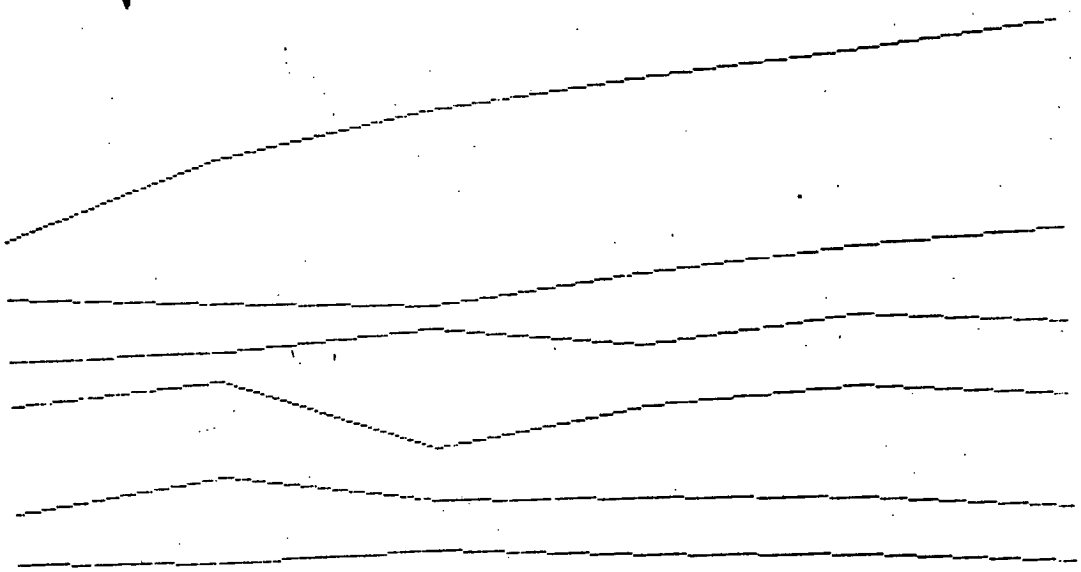
Evolution des candidatures au bac (Acad. de NANTES)

TRAITEMENT

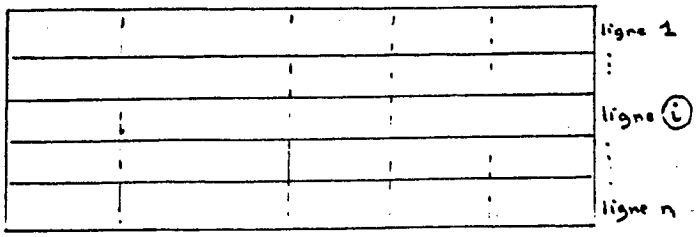
partie STATIQUE



partie DYNAMIQUE débouchant sur

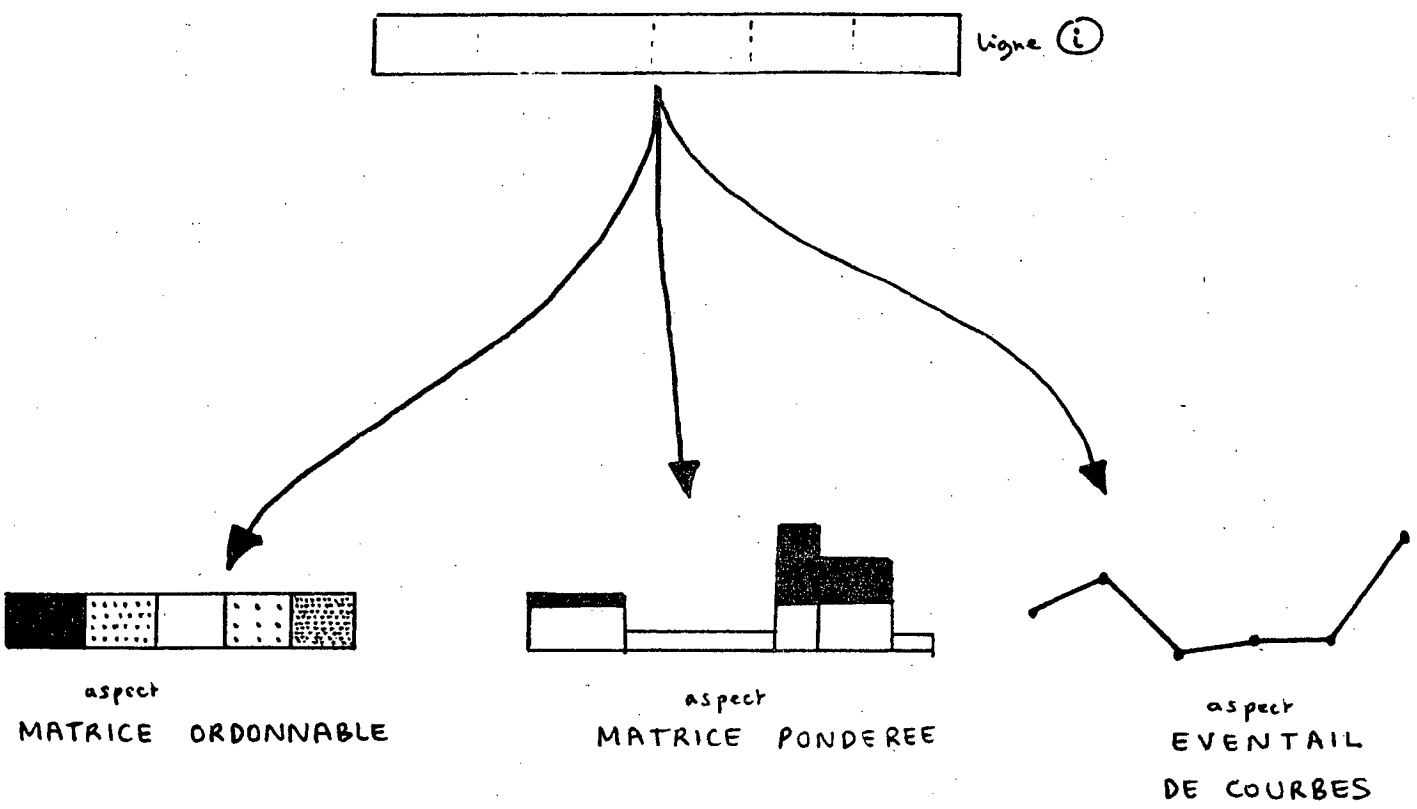


ANALOGIE



graphique matriciel Bertin
issu d'un TRAITEMENT QUELCONQUE
dans un ETAT QUELCONQUE

DIFFERENCES



Le traitement est décomposé en ses deux parties :

- partie **statique** : le « **décalque** » des données se fait en les représentant par des surfaces proportionnelles à leur poids, la partie noircie éventuelle correspond à un poids élémentaire supérieur au poids de la ligne (profil) ;
- partie **dynamique** :
  - un profil « marqué » est choisi (ici celui du haut) ; toutes les colonnes sont reclassées selon le classement décroissant de ce profil ; le profil de référence est replacé en haut (ici il y était déjà) ; cette opération « complexe » revient à rapprocher les colonnes **se ressemblant visuellement** (voisines selon une certaine distance visuelle d) ;
  - les profils (lignes) sont reclassés : les plus **proches visuellement** sont rassemblés, les plus différents éloignés.

3. Le Document 16 présente schématiquement un traitement en **éventail de courbes.**

Le tableau de données indique l'évolution du nombre de candidatures au baccalauréat dans l'académie de Nantes entre 1976 et 1981, dans les séries A,B,C,D,D' et E [ACAD81].

Le traitement est décomposé en ses deux parties :

- partie **statique** : le « **décalque** » des données se fait en associant à chaque série (ligne) une courbe ;
- partie **dynamique** : elle se limite au rapprochement des **lignes** (courbes) **se ressemblant visuellement** puisqu'en optant pour ce traitement on a admis que la composante colonnes (années) était ordonnée à priori.

Une fois le graphique final obtenu, on peut faire apparaître les noms des séries et des années pour passer à l'interprétation.

3. Schéma : analogie des traitements/ différences

Schématiquement, nous pouvons considérer :

- que tous les traitements sont **formellement analogues,**
- les **différences réelles** n'apparaissant que dans le **type d'objets élémentaires constituant les lignes**.

Le Document 17 schématise cela.

[Notons que si trois exemples seulement ont présentés, c'est parce qu'ils suffisent à recouvrir tous les traitements Bertin vus sous l'angle représentation graphique des données].

Précisons ces différences :

- représentation type « **matrice ordonnable** » : les objets élémentaires sont des **rectangles** de **tailles identiques** dont la teinte (≡ intensité de noir) est fonction de la valeur représentée ;
- représentation type « **matrice pondérée** »: les objets élémentaires sont des **rectangles** de **surfaces variables** (longueur et largeur variable) proportionnelles au poids de la donnée représentée ;
- représentation type « **éventail de courbes** » : les objets élémentaires sont des **segments** de droite dont la projection sur l'axe horizontal est de longueur constante mais **dont la pente varie** en fonction des valeurs des deux données qu'ils « joignent ».

4. Conséquences « pédagogiques »

En tenant compte du traitement demandé par l'utilisateur, du type de représentation graphique adopté, de la nature des données (binaires ou non), nous pouvons dénombrer **onze traitements matriciels ou variantes distinctes**.

Cependant, compte tenu des nalogies entre ces traitements, mises en évidences précédemment et reposant sur le concept de traitement matriciel, l'utilisateur n'aura pas onze apprentissages distincts à effectuer. **Un seul apprentissage,** bien mené, surtout si c'est celui du traitement de base, la matrice ordonnable, **suffira pour l'essentiel : s'imprégner de l'« esprit Bertin ».**

Toutefois, pour qu'il en soit effectivement ainsi, l'organisation du logiciel doit être adaptée. Alors, l'apprentissage de l'utilisation du logiciel sera considérablement simplifiée : ce sera aussi un atout pour un « développement » ultérieur.

5. Conséquences informatiques

1. Tous les traitements doivent donc être **présentés de façon analogue à l'utilisateur : l'organisation du dialogue** doit donc être basée sur leur analogie (pour la solution apportée, voir VI et VII).

2. La mise en évidence de la grande analogie des traitements associée à une perception claire de leurs différences incite à la création d'un **vaste « jeu » de procédures** (modules) **minuscules** ; chaque procédure sera utilisable par plusieurs traitements et variantes.

Alors, une fois un traitement choisi par l'utilisateur et resitué par le système parmi les onze (on peut toujours multiplier les variantes) traitements répertoriés, le système pourra automatiquement établir le « **contexte** » du traitement. Il établira la **liste des procédures nécessaires** (connue à priori) et adoptera une **stratégie pour leur chargement** en mémoire centrale : cette liste peut être établie à partir d'un tableau de booléens à deux dimensions (traitement, procédure).

Cette façon de procéder (établir le « contexte » **en puisant** dans le « jeu » de procédures) a, outre des avantages de programmation, un intérêt certain pour le **développement du logiciel sur micro-ordinateur** :elle permet une utilisation rigoureuse, sinon optimale, des possibilités de **mémoire**.

VI. UN LOGICIEL STRUCTURE SELON LES CHOIX DE L'UTILISATEUR

1. Priorité à l'utilisateur/ priorité au dialogue

Le logiciel est un logiciel d'**AIDE** au traitement des données, non pas de traitement tout court. Ce qu'il faut viser, c'est comme le dit Bertin d'« **aller au-delà de ce qui est automatisable** ». En effet : « on trouvera de nombreux moyens d'automatiser la diagonalisation d'une matrice. On ne trouvera jamais le moyen d'automatiser l'imagination des sous-ensembles les plus utiles. **Les manipulations graphiques ne peuvent être menées à leur terme que par le « décideur » lui-même** » ([BER77] p.21/je souligne).

L'organisation du logiciel doit donc refléter la suprématie de l'utilisateur sur le système qui n'est que l'exécutant : elle **doit être décrite par la structure du dialogue.**

2. La hiérarchie fondamentale et son expression formelle (voir Document 18)

Exprimer la hiérarchie [utilisateur→système] implique de bien séparer leurs rôles puis de définir avec précision les points d'articulation entre ces rôles : quand les ordres sont-ils donnés et exécutés ?.

Une expression formelle simple peut être donnée par une **grammaire** (voir VII). Il ne s'agit pas là simplement d'un exercice de style mais de définir un **outil puissant d'écriture du logiciel**...et donc de réécriture et de modifications éventuelles (voir IX).

HIERARCHIE FONDAMENTALE

Choix — niveau Utilisateur non automatique

↓

EXECUTION DES CHOIX — niveau Système automatique

HIERARCHIE DES CHOIX

Niveau ⓪ : concerne la globalité du graphique

Niveau ① : concerne une partie du graphique

↓

Niveau ②

↓

Niveau ③

↓

.....

globalité ?

UTILITE croissante

FACILITE D'ACCES croissante

HIERARCHIE D'EXECUTION

EXECUTION NUMERIQUE — effecte la copie numérique du graphique

↓

EXECUTION GRAPHIQUE — effecte le graphique

Dans cette grammaire, l'organisation des **choix** de l'utilisateur (niveau supérieur de la hiérarchie) est exprimée par les « **règles de production** », **l'exécution** de ces choix par le système par des « **actions sémantiques** », **les points d'articulation** par les **lieux d'insertion** de ces actions.

Puisque donner la priorité à l'utilisateur c'est le dispenser des tâches laborieuses pour favoriser son rôle d'acteur du traitement, le logiciel sera conçu pour **ne pas l'accabler de questions redondantes.** Pour cela, le système mémorisera certaines réponses afin de déduire, le moment venu, certains choix, sans solliciter l'utilisateur. Cependant, de tels choix doivent être exprimés au niveau supérieur : un formalisme adapté permettra donc **d'exprimer des choix faits par le système** (comme **conséquences de choix antérieurs de l'utilisateur**, répétons-le) au sein de la grammaire (voir VII.2.).

Dans le même ordre d'idées, le système pourra **déterminer dynamiquement** (en cours d'exécution) **certains éventails de choix**. Un formalisme adapté est proposé (voir VII.2.).

3. L'organisation des choix (niveau supérieur)

1. **L'analogie** fondamentale **entre les traitements** matriciels mise en valeur précédemment, la nécessité « pédagogique » de présenter ces traitements de façon analogue à l'utilisateur, conduisent à **écrire le logiciel** i.e. les règles de production de la grammaire **sous une forme indépendante du traitement envisagé** et simplement **paramétrée**.

Les règles de production « effectives » se déduiront des règles « formelles », indépendantes du traitement, selon la valeur du paramètre (voir VII).

On pourra ainsi très rapidement prendre connaissance des analogies, des différences, ajouter au traitement i_1 telle possibilité présente dans le traitement i_2, etc...

2. Aider l'utilisateur, c'est lui offrir à chaque étape les possibilités les plus riches...sans l'étouffer ou le désorienter par une gamme trop large. D'où une **hiérarchisation des choix** proposés dans la partie dynamique du traitement (voir Document 18).

3. Notons enfin que **l'organisation formelle des gammes de choix** (grammaire) est tout à fait **indépendante** de la **réalisation effective des choix** (façon dont le système interroge l'utilisateur, dont ce dernier lui répond). Cette réalisation est par contre très liée au **matériel** sur lequel le logiciel peut être implanté : dispose-t-on d'un sélecteur de menu, d'un photostyle, peut-on superposer modes graphique et non graphique, etc ?. Cette réalisation peut aussi varier selon le traitement ou le point atteint dans un traitement particulier.

4. L'organisation de l'exécution des choix (niveau inférieur)

1. L'exécution des choix de l'utilisateur par le système débouche généralement sur une modification du graphique affiché.

Tenir compte de l'analogie entre les traitements d'une part, vouloir isoler les instructions graphiques pour faire mieux apparaître les primitives essentielles d'autre part, amènent à **hiérarchiser l'exécution des choix**.

2. **Le niveau « exécution numérique »** sera directement commandé par le niveau supérieur et commandera pour sa part **le niveau « exécution graphique »** (voir Document 18).

3. **L'exécution numérique** assurera la **mise à jour de la « copie numérique » du graphique** (copie « comprimée » bien sûr et non copie-mémoire de l'écran!) en fonction des choix de l'utilisateur. Pour l'essentiel, elle sera **indépendante du traitement** particulier en cours ; les actions sémantiques d'exécution numérique seront elles aussi paramétrées (voir VIII.2.).

4. **L'exécution graphique** assurera les **modifications d'écran** découlant de l'exécution numérique (voir VIII.3.). Hormis ce qui touche à la réalisation effective des choix (cf. Supra), le niveau « exécution graphique » sera le seul à contenir des instructions graphiques. D'une façon globale cette fois, ce sera le seul à pouvoir modifier durablement l'écran ; toute opération de dialogue affectant l'écran ne le fait que momentanément (éventuellement à travers : sauvegarde - restitution).

VII. DESCRIPTION DU LOGICIEL ET COMMENTAIRES

1. Introduction

Voici la description détaillée du logiciel (*) sous forme de grammaire. Des commentaires accompagnent chaque étape.

Mode d'emploi : le logiciel est décrit sous forme de fiches : fiches « grammaire » G, fiches « commentaires » C, numérotées. Il faut consulter les fiches G et C de numéros identiques.

2. Formalisme adopté

La grammaire est décrite par des règles de production. On suppose que lors du choix parmi plusieurs alternatives, il convient de s'arrêter à la première satisfaite en partant de la gauche.

Le vocabulaire terminal est souligné : -d'un trait simple, s'il appartient à une gamme de choix définie statiquement (ou en caractères gras dans G6 et G7) -d'un trait double s'il appartient à une gamme créée dynamiquement.

Les choix effectués par le système sont marqués d'une astérisque.

Quelques actions sémantiques simplement, ont été insérées, pour illustrer sans encombrer.

(*): dans sa partie « traitement graphique » mais incluant le choix d'un traitement graphique particulier.

3. Description

G1 - Avant le traitement proprement dit

(R1) ⟨traitement graphique⟩ ::= traitement graphique ⟨tentative de traitement⟩ arrêt du système*

(R2) ⟨tentative de traitement⟩ ::= abandon ⟨liste des noms de tableaux disponibles⟩ s_2 ⟨traitement⟩

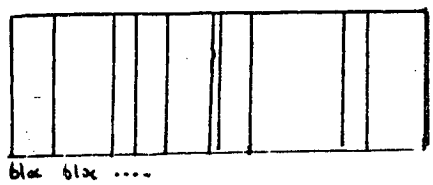
(R3) ⟨liste des noms de tableaux disponibles⟩ ::= nom 1 | nom 2 | ... | nom t

(R4) ⟨traitement⟩ ::= s_{41} ⟨choix du traitement graphique⟩ s_{42} ⟨corps du traitement graphique⟩

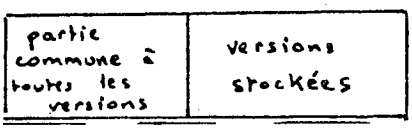
(R5): ⟨corps du traitement graphique⟩ ::=

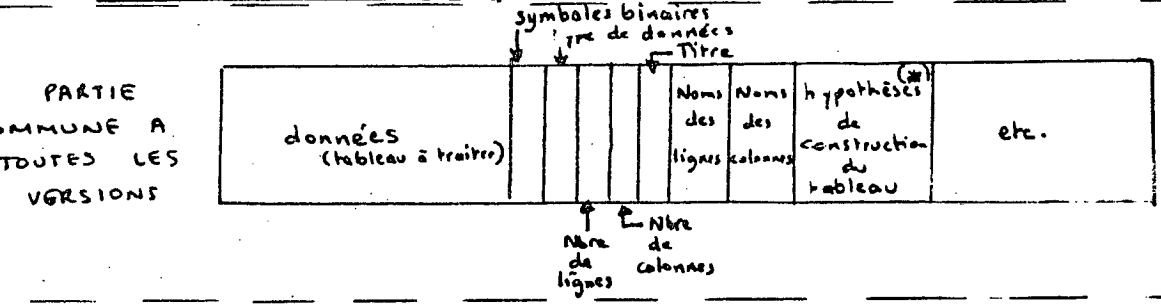
- [ordonnable, ordonnable, binaire]* s_{51} ⟨T_1⟩ s_{52}
- [ordonnable, ordonnable, non binaire]* ⟨T_2⟩
- [ordonnable, pondérée, binaire]* ⟨T_3⟩
- [ordonnable, pondérée, non-binaire]* ⟨T_4⟩
- [ordonnable, fichier matrice, binaire]* ⟨T_5⟩
- [ordonnable, fichier matrice, non-binaire]* ⟨T_6⟩
- [ordonnable, fichier image, binaire]* ⟨T_7⟩
- [ordonnable, fichier image, non binaire]* ⟨T_8⟩
- [pondérée, pondérée, quelconque]* ⟨T_9⟩
- [courbes, éventail, rapport]* ⟨T_{10}⟩
- [courbes, éventail, différence]* ⟨T_{11}⟩
- [?, non matriciel - plus de 3 caractères, ?]* ⟨T_{12}⟩
- [?, 3 caractères au plus, ?]* ⟨T_{13}⟩

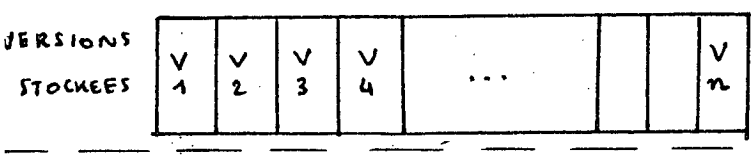
ORGANISATION DE LA MEMOIRE DE TRAITEMENT

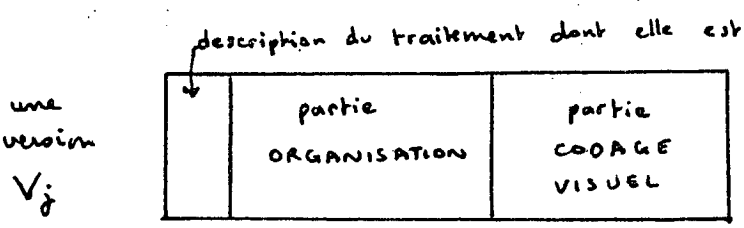


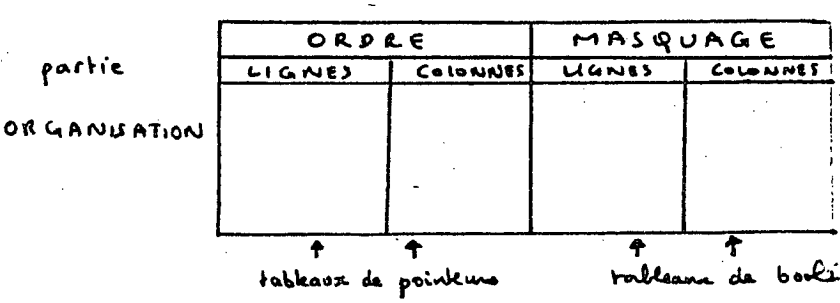
chaque bloc est associé à un tableau "traitable" ; il le contient

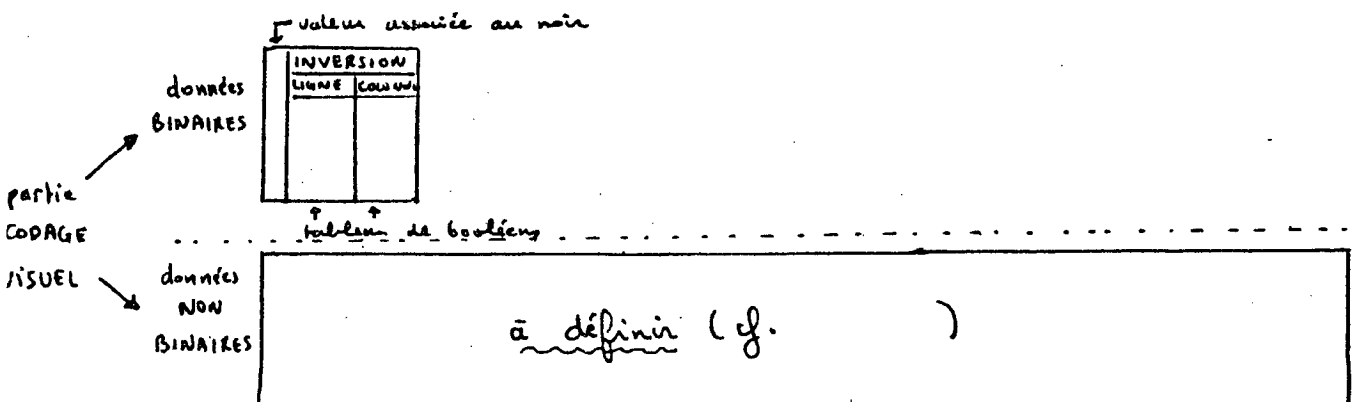












(*) hypothèses de construction du tableau : peuvent servir pour [choix du traitement], [définition du codage visuel], [réalisation du document de travail].

LES DIFFERENTS TYPES DE REPRESENTATION GRAPHIQUE(*) ASSOCIES AUX TRAITEMENTS MATRICIELS (**)

TRAITEMENT MATRICIEL / REPRESENTATION GRAPHIQUE	MATRICE ORDONNABLE	FICHIER MATRICE	FICHIER IMAGE	MATRICE PONDEREE	EVENTAIL DE COURBES
MATRICE ORDONNABLE	toujours	toujours	toujours	si tableau "trop grand": règles spécifiques de choix des paliers	
MATRICE PONDEREE				si "petit" tableau	
EVENTAIL DE COURBES					toujours

(*) types de représentation graphique : voir
(**) traitements matriciels : voir

LES DIFFERENTS TRAITEMENTS ET VARIANTES (*)

classés en fonction de :

- type de traitement matriciel choisi
- type de représentation graphique résultante
- nature ou type des données
  - nature : cf. Eventail de courbes : différences ou rapports significatifs ?
  - type : linéaire ou non ?

(*) cf. (G1)(R5).

C1

.(R1): l'utilisateur a choisi d'effectuer un traitement graphique, il veut le tenter ; il pourra y renoncer s'il s'aperçoit (cf. (R3): liste créée dynamiquement par le système) qu'aucun tableau lui convenant n'est présent en « mémoire de traitement » (cf. Document 19): « **abandon** » (règle (R2)).

.s_2 cette action sémantique enregistre le choix d'un tableau précis.

.(R4): l'utilisateur va choisir un traitement graphique : on suppose qu'il peut en trouver un, adapté, car les tableaux chargeables en mémoire de traitement ont été sélectionnés (i.e. ne sont pas trop grands) pour supporter au moins un type de traitement graphique.

.s_{41} s_{42} ces actions définissent un bloc où seront déclarées les variables nécessaires au choix du traitement.

.(R5): en fonction du choix effectué en (R4) par l'utilisateur, le système, qui a valorisé l'un des treize triplets, peut choisir parmi les treize alternatives. Ces triplets définissent un traitement graphique (ou une variante) précis : signification des différentes valeurs (voir Document 20):

- valeur 1 : qualifie le type de représentation graphique (cf.Document 20,V.3.);
- valeur 2 : qualifie le type de traitement choisi;
- valeur 3 : qualifie la nature des données;

la définition n'est pas complète pour les deux derniers traitements (?) car ils n'ont pas encore été étudiés.

.Les s_{51i} et s_{52i} encadrent le bloc dans lequel le traitement sera effectué ; l'exécution des s_{51i} génère le contexte propre au traitement (cf. V.5.) : ces actions n'ont été notées (pour ne pas alourdir) que pour i = 1.

G2 - Choix d'un traitement graphique

(R1) <Choix du traitement graphique> :: = s_1 <choix si plus de 3 caractères> s_1 <choix si 3 caractères au plus>
(R2) <Choix si plus de 3 caractères> :: = <il y a plus de 3 caractères> <suite du choix si plus de 3 caractères>
(R3 <Il y a plus de 3 caractères> :: = <plus de 3 lignes et plus de 3 colonnes>* <la composante caractères a plus de 3 éléments>
(R4) <La composante caractères a plus de 3 éléments> :: = plus de 3 lignes et moins de 3 colonnes* les lignes sont les caractères | plus de 3 colonnes et moins de 3 lignes* les colonnes sont les caractères
(R5) <suite du choix si plus de 3 caractères> :: = vouloir construction matricielle <choix du traitement matriciel> | ne pas vouloir de construction matricielle <choix de traitement non matriciel>
(R6) <choix du traitement matriciel> :: = il y a une composante déjà ordonnée <choix de composante> | <choix de traitement à une composante déjà ordonnée> | aucune composante n'est déjà ordonnée <choix de traitement à deux composantes ordonnables>
(R7) <choix de composante> :: = ligne | colonne
(R8) <choix de traitement à une composante déjà ordonnée> :: = fichier-image | <demande d'éventail de courbes>
(R9) <demande d'éventail de courbes> :: = différences significatives <les rapports peuvent-ils être significatifs?>
(R10) <les rapports peuvent-ils être significatifs ?> :: = au moins un zéro* | aucun zéro* rapports significatifs
(R11) <choix de traitement à deux composantes ordonnables> :: = matrice ordonnable | fichier-matrice <suite si fichier matrice demandé> | matrice pondérée <suite si matrice pondérée demandée>
(R12) <suite si fichier matrice demandé> :: = <choix de composante>
(R13) <suite si matrice pondérée demandée> :: = matrice de comptage | non
(R14) <choix de traitement non matriciel> :: = ...
(R15) <choix si 3 caractères au plus> :: = ...

C2

.Il s'agit là simplement de la réécriture du « **synoptique** » proposé par Bertin ([BER77] p.28, 101, 233).

.(R1) : orientation soit vers la **graphique « dynamique »** soit vers la **graphique « statique »** ; grâce à s_1, le système prend connaissance es dimensions du tableau.

.Remarquer **l'imbrication des choix effectués par le système** pour décharger l'utilisateur **et de** ceux qui sont faits directement par ce dernier ; la règle (R4) est un bon exemple : - c'est le système qui répond quant à la taille du tableau (il l'a apprise en s_1 grâce à l'information donnée précédemment par l'utilisateur : le nom du tableau à traiter) - c'est l'utilisateur qui indique ce qui est « objet », ce qui est « caractère » (ça peut changer d'un traitement à l'autre suivant la façon de poser le problème).

.(R5) : bifurcation entre **constructions matricielles** ou non.

.(R6) : voilà l'exemple même d'une règle réutilisable...donc potentiellement d'une des procédures minuscules évoquées précédemment.

.**C'est toujours l'utilisatuer qui donne leur sens aux données :** en disant ce qui est « caractère », « objet », si une composante est ordonnée a priori, si les différences ou rapports entre les données d'une même ligne ont un sens, si la matrice est « de comptage » (cf.[BER77]).

G3 - Structure du traitement

(R1) $\forall i \in \{1,\ldots,12\}$:
T_i ::= <phase statique i><phase dynamique i><précaution>
si i = 13 :
T_i ::= <phase statique i> <fin traitement>

(R2) <précaution> ::= **stockage**<stocker> | **non**

(R3) <fin traitement> ::= **rien** | **document** s_3

C3

.Cette fiche exprime **l'opposition entre graphique « dynamique »** (i < 13) où le graphique initial sera manipulé **et graphique « statique »** où le traitement est terminé de fait dès que le graphique initial (et final) est obtenu (i = 13).

.(R2) : mettre en garde l'utilisateur contre une destruction involontaire de son travail : lui permettre le stockage in.extremis de l'état graphique atteint et des codages visuels effectués.

(R3).s_3 : l'utilisateur pourra obtenir la copie de l'écran contenant le graphique « classique » demandé.

G4 - Phase statique du traitement

(R1) $\forall i \in \{1,\ldots,13\}$:
<phase statique i> ::= s_{11} <information de base i> s_{12i} s_{13i}

(R2) $\forall i \neq 13$:
<information de base i> ::= standard <fin standard i> <tentative de reprise i>

(R3) Si $i \in \{1,5,7\}$: <fin standard i> ::= valeur binaire 0 associée au noir | valeur binaire 1 associée au noir
Si $i \in \{2,8\}$: <fin standard> ::= <choix de composante>

(R4) $\forall i \neq 13$:
<tentative de reprise i> ::= il y a version stockée* <choix de reprise i> | il n'y a pas de version*

(R5) $\forall i \in \{12,13\}$:
<choix de reprise i> ::= globale <reprise globale i> | patchwork <reprise patchwork i>

(R6) $\forall i \in \{12,13\}$:
<reprise globale i> ::= abandon | <liste des noms de versions disponibles>

(R7) $\forall i \in \{12,13\}$:
<reprise patchwork i> ::= <reprise ordre i><reprise masquage i><reprise codage visuel i>

(R8) $\forall i \in \{12,13\}$:
<reprise ordre i> ::= abandon | <liste des noms de versions disponibles>

ETC...

C4 - Phase statique du traitement

- La phase statique du traitement aboutit toujours à la **génération du graphique initial** (action s_{13i}).
- Elle inclut toujours la **collecte de l'« information de base »** nécessaire à cette génération et à la partie dynamique ((R1)).
- La génération u graphique initial, une fois l'information de base collectée, impose la définition par le système des **paramètres de visualisation** (action s_{12i}) :

- définir la **taille des matrices élémentaires** (représentation en matrice ordonnable), **centrer le graphique** de façon à occuper au mieux l'écran ; les matrices élémentaires doivent être assez grandes pour ne pas fatiguer la vue de l'utilisateur, pas trop grandes pour lui conserver une vision globale sans lui imposer de reculer à dix mètres de l'écran !,
- définir la **palette des niveaux visuels** (représentation en matrice ordonnable);
- définir l'**échelle** du graphique et **l'espacement** entre les profils ou les courbes (représentation en matrice pondérée ou éventail de courbes).

- La collecte de l'information de base passe par un **accès aux données** à traiter : le système l'assure lui-même (en fonction du tableau indiqué par l'utilisateur (cf. G1,(R2),(R3))) par l'action s_{11i}. Cet accès n'a lieu que dans le bloc défini par ⟨phase statique i⟩, sauf à autoriser le recodage visuel en cours de traitement (recodage dynamique : voir X.4.). C'est important pour la gestion de la mémoire ; une **version simplifiée du logiciel**, sans redéfinition dynamique des codages visuels, pourra se contenter d'une capacité mémoire très limitée.
- Si $i \neq 13$: **la collecte de l'information de base** est toujours réalisée ((R2)) :

- **soit de façon « standard »** : ordre des lignes et colonnes conforme à l'ordre de stockage du tableau dans la mémoire de traitement, aucune ligne ni colonne n'est masquée, les « choix » de codage visuel sont « standard » quand les données sont non binaires et que de tels « choix » s'imposent (voir Document 21) ;
- **soit comme « reprise d'un traitement antérieur »** : règles (R4) et suivantes.

- Quand l'utilisateur veut reprendre un traitement antérieur, le système analyse lui-même si ce traitement est **compatible** avec le traitement actuel et sous quel(s) aspect(s) ; il en trie les conclusions nécessaires pour définir les gammes de choix offertes à l'utilisateur.
- La reprise peut s'effectuer :

- **globalement** : reprise globale d'un « état graphique » antérieurement stocké ;
- **en « patchwork »** i.e. en ne reprenant que certains aspects d'« états graphiques » antérieurs et en les combinant.

- La possibilité de reprise d'un traitement est **essentielle** car elle permet à l'utilisateur de suspendre un traitement pour réfléchir, **interpréter** : elle est aussi fondamentale pour permettre de tenter **plusieurs orientations de traitement** ; en ce sens, la possibilité de « patchwork » multiplie les orientations envisageables (voir III.6.).
- La connaissance de l'« **organisation** » du graphique (ordre des lignes et colonnes + masquage) et de codage visuel adopté (d'où la « **copie codée** » des données) permet au système de disposer de la « **copie numérique** » **du graphique** (voir VIII.2.) à partir de laquelle est généré le graphique initial.

G5 - Le niveau 0 de la phase dynamique

(R1) $\forall i < 13$: ⟨phase dynamique i⟩ ::= ⟨niveau 0 i⟩

(R2) $\forall i < 13$: ⟨niveau 0 i⟩ ::= ⟨choix niveau 0 i⟩ tq(¬(fin du traitement))* ⟨action selon choix niveau 0 i⟩⟨choix niveau 0 i⟩ftq*

(R3) $\forall i < 13$:

⟨choix niveau 0 i⟩ ::= stockage | document | manipulation | fin du traitement

on pourra ajouter par exemple :

Si $i \in \{3,4,9\}$: reclasser selon profil.ligne

Si $i \in \{1,2,3,4,7,8\}$: modification globale des choix de codage visuel

Si $i = 9$: reclasser globalement les profils

etc.

(R4) $\forall i < 13$

⟨action selon choix niveau 0 i⟩ ::= ⟨stocker i⟩* ⟨documenter i⟩⟨niveau 1 i⟩*

où on ajoutera par exemple (cf.supra) :

Si $i \in \{3,4,9\}$: ⟨choisir un élément de composante i⟩

Si $i \in \{1,2,3,4,7,8\}$: ....

Si $i = 9$: ⟨liste ordonnée des profils i⟩*

etc.

ETC

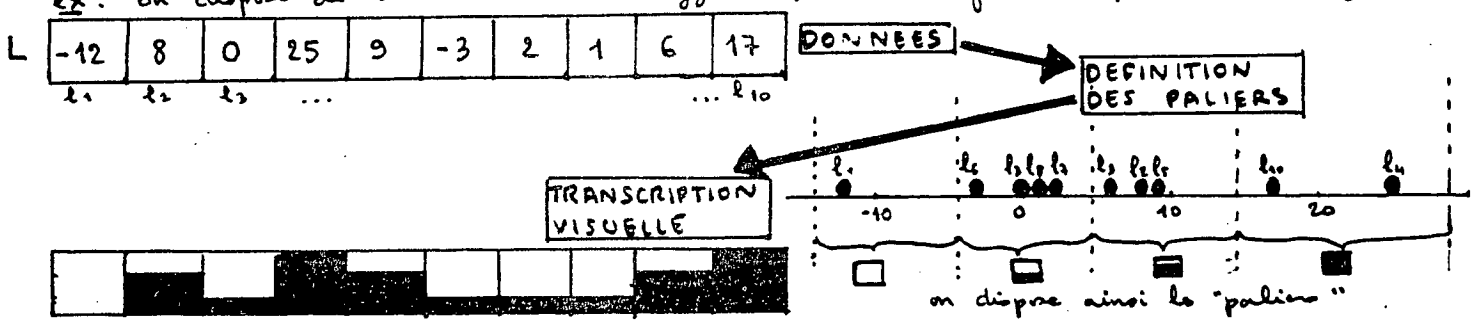
MATRICE ORDONNABLE ou FICHIER-IMAGE

CODAGE VISUEL STANDARD

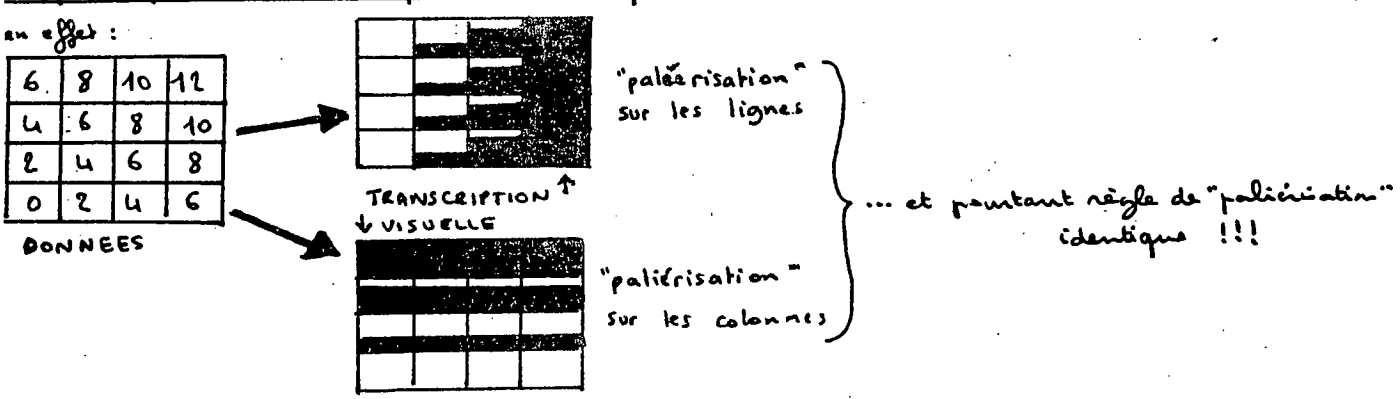
DONNEES NON BINAIRES

il faut définir des "paliers" :

ex : on dispose de 4 niveaux visuels différents, on veut définir des paliers sur la ligne L



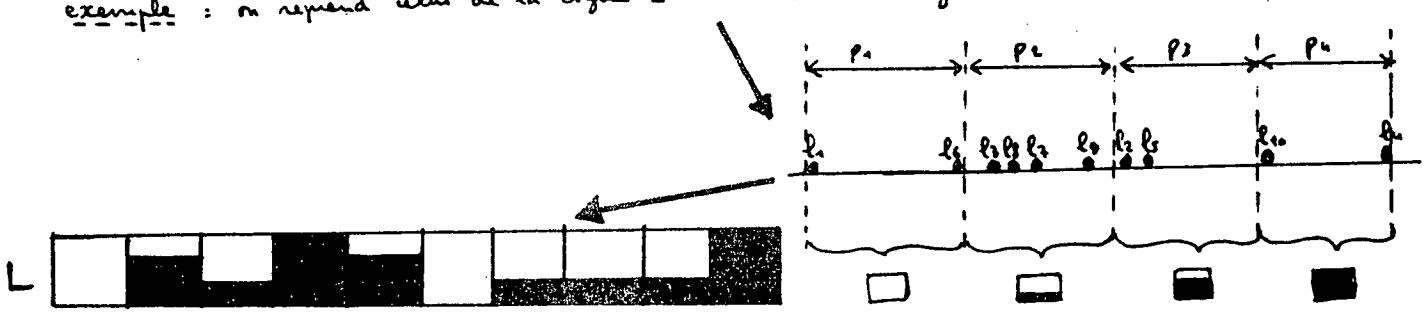
il faut préciser la composante de "paliérisation" :



CODAGE STANDARD

- l'utilisateur indique seulement la composante de "paliérisation" (cf. (G4)(R3))
- le nombre de niveaux visuels est fixé par le système selon une option "standard"
- la "paliérisation" est "équidistante" et part des extrema inclus

exemple : on reprend celui de la ligne L ci-dessous avec toujours les 4 mêmes niveaux visuels :



$p_1 = p_2 = p_3 = p_4$ $\quad l_4 - l_1 = p_1 + p_2 + p_3 + p_4$

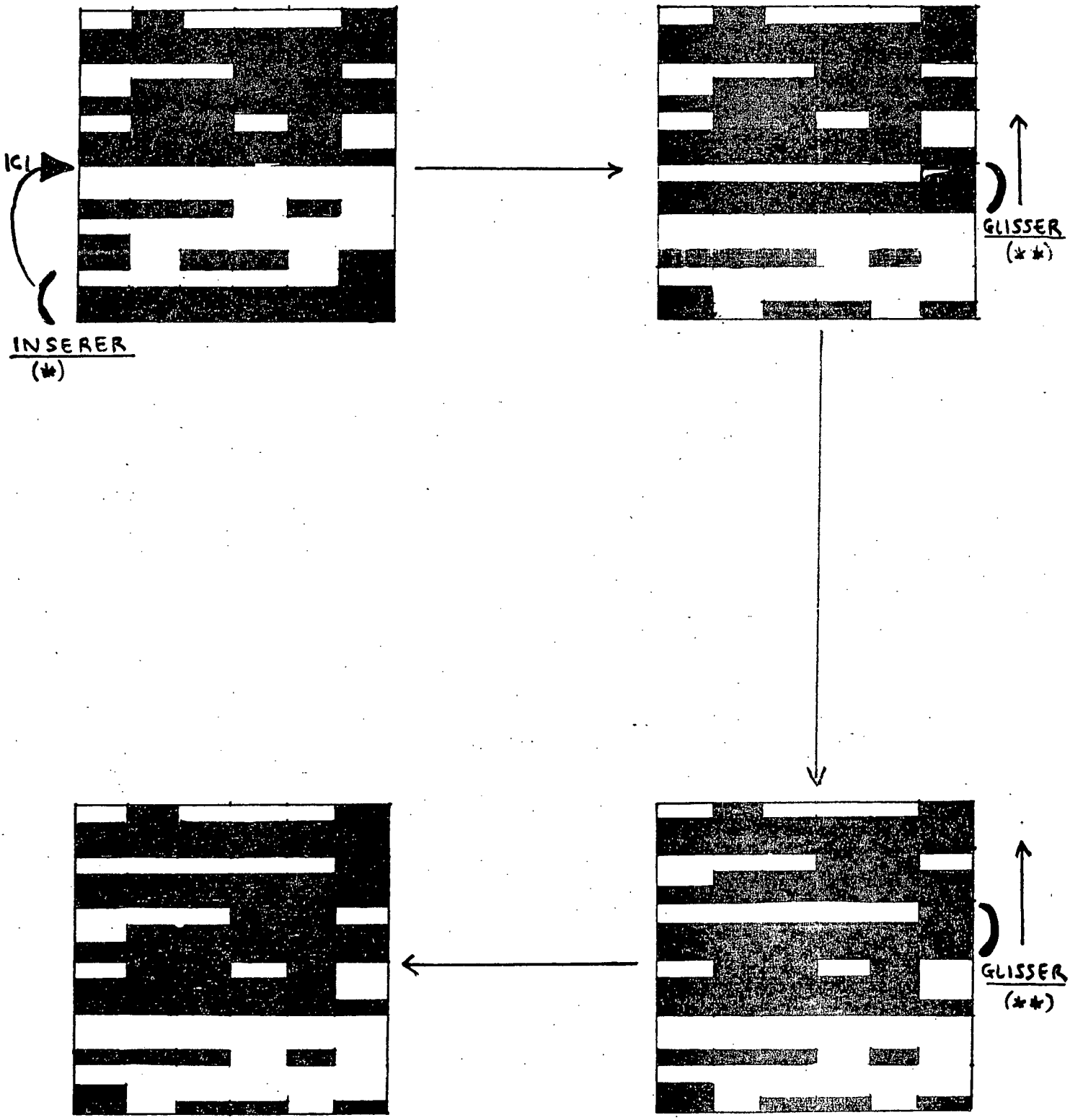
C5 - Le niveau 0 de la phase dynamique

- L'organisation des choix de la phase dynamique est **hiérarchique** (voir VI.3.). Les niveaux 0 et 1, les plus accessibles, sont les lieux des commandes les plus **utiles**. Le niveau 0 réunit les commandes très utiles agissant **globalement** : leur utilité n'implique pas un emploi très fréquent contrairement au niveau 1 où les commandes agissent sur des parties souvent élémentaires du graphique.
- D'autres opérations globales peuvent exister dans la partie dynamique, renvoyées aux niveaux 2,3,...: on peut s'en dispenser : la plupart des versions du logiciel n'en seront pas pourvues. Ainsi, seules les opérations de réclassement automatique (global) fondamentales seront disponibles au niveau 0 pour les traitements où il en existe (ex : matrice pondérée, fichier-matrice) (voir X.2.).
- Remarquer que l'utilisateur peut se maintenir au **niveau 0 aussi longtemps qu'il le désire** (boucle TQ, évaluée par le système) ; chaque niveau est organisé ainsi.
- Au niveau 0 seront toujours disponibles **deux opérations essentielles : stocker** l'état graphique obtenu et les choix de codage effectués + obtenir un **document** de travail conforme à l'état graphique obtenu ; ce sont là deux exemples d'opérations globales dont l'exécution numérique n'entraîne pas d'exécution graphique (voir VIII.).
- **Le DOCUMENT de TRAVAIL** :

- il est créé **automatiquement** (car aucun choix de mise en page : ce n'est pas un « graphique de communication ») à la demande ;
- il doit **aider l'utilisateur** à poursuivre le traitement et/ou à interpréter les résultats obtenus ;
- il comporte donc **deux volets** : la « copie nue » du graphique (pas de nom d'objet ni de caractère) et la « copie habillée » (noms d'objets, de caractères, choix de codage visuel effectués) ;
- la « **copie nue** » du graphique peut permettre d'anticiper le traitement en envisageant mentalement de **nouveaux rapprochements visuels ;**
- la **« copie habillée »** permet d'interpréter les résultats obtenus (**aide à l'interprétation** : voir III) et donc de choisir la suite à donner (voir III dont III.6.) : c'est pourquoi elle contient les informations qui permettent de resituer le traitement dans le « **monde extérieur** » : noms des objets, caractères, choix de codage visuel.
- quand les données sont non binaires et que des « paliers » ont été définis pour les coder, le document pourra présenter ainsi ces **choix de codage** :

. valeurs à coder,
. diagramme de répartition + place des paliers,
. transcription graphique correspondante (niveaux de grisé),
. fonction de « paliérisation » (voir X.4.).

- Le **STOCKAGE** de l'état graphique courant : quand l'utilisateur veut **suspendre un traitement** pour réfléchir, interpréter, il doit pouvoir disposer d'un délai indéfini...tout en libérant la machine pour d'autres utilisations. Mais il ne doit pas être « pénalisé » et doit donc pouvoir ré-obtenir automatiquement (i.e. rapidement, sans fatigue, sans erreur) l'état graphique précédemment atteint s'il veut le ré-utiliser (voir III et X). D'où la possibilité de stocker cet état graphique. Par la même occasion, cela permettra à l'utilisateur de **tenter plusieurs directions de traitement** (voir II.6.). Il sera donc amené à stocker les états graphiques qui lui paraissent intéressants même quand il n'envisage pas de suspension du traitement ; il le fera notamment quand il sent qu'il opère une ré-orientation décisive (ex : modification globale des choix visuels) afin de pouvoir, le cas échéant, revenir aisément en arrière.
- Modification globale des choix visuels (voir X.4.) : il s'agit par exemple : -d'inverser noir et blanc globalement quand les données sont binaires -d'écarter globalement un objet ou un caractère pour redéfinir les « paliers » quand les données sont non binaires, de changer le nombre de niveaux visuels disponibles (varier le contraste)- ; notons que cette modification globale du codage visuel implique l'**accès global aux données** à traiter (cf. X.4.).

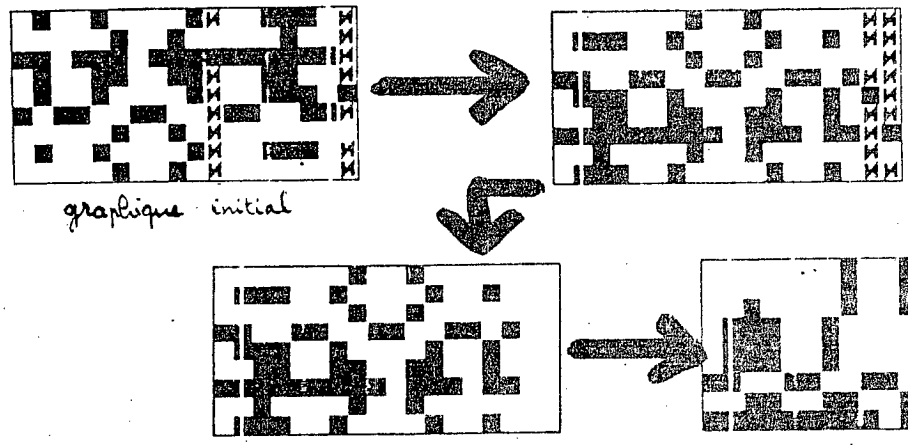


(*) INSERTION : on "dégrossit"

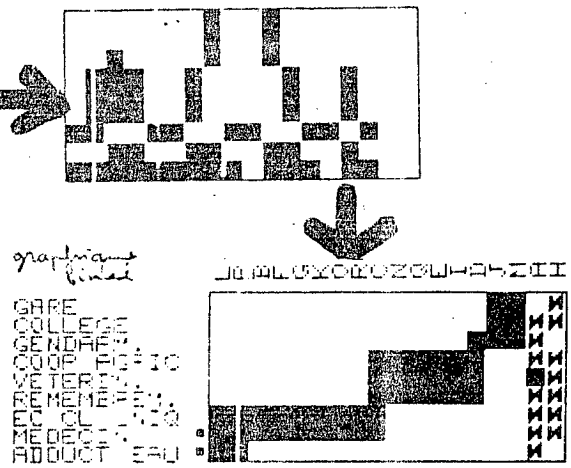
(**) GLISSEMENT : on "affine"

MASQUAGE DE LIGNES ET/OU COLONNES

présentation et exemple 1



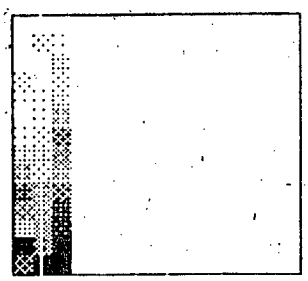
graphique initial

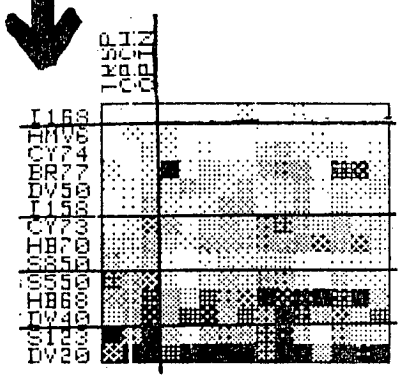


ne pas être gêné par le "bruit" dû aux données manquantes

les éléments masqués sont cependant entraînés dans les manipulations par le système

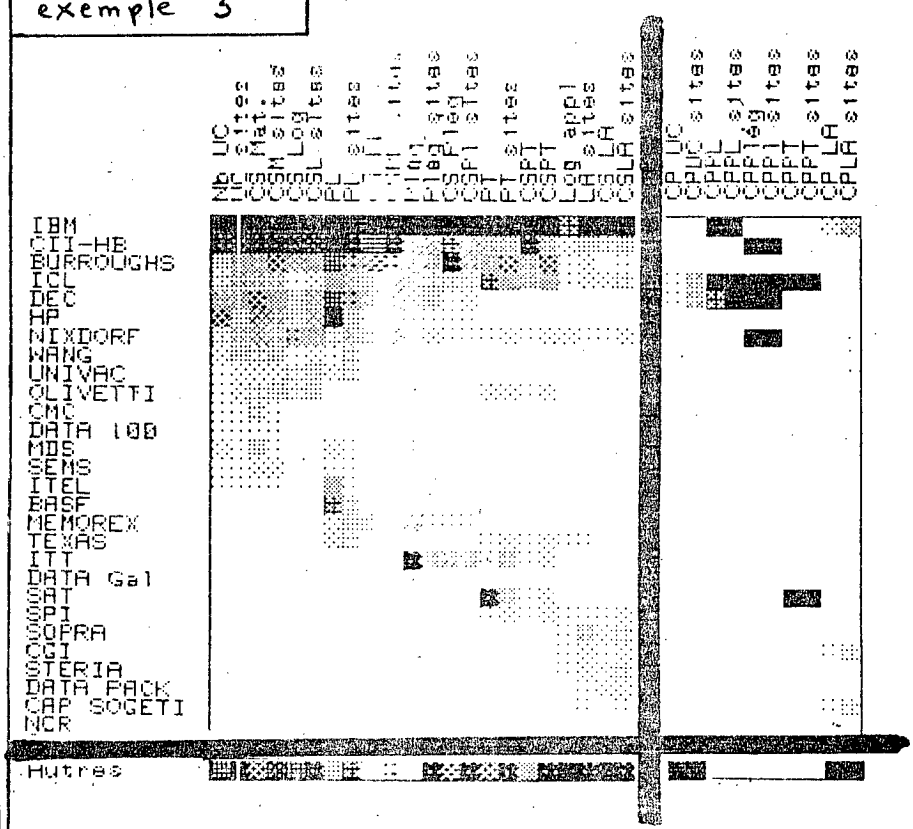
exemple 2





analyser selon 3 caractères

exemple 3



mettre en évidence plusieurs systèmes de données

DOCUMENT 23

G6 - Le niveau 1 de la partie dynamique

Il n'est présenté, **à titre d'exemple,** que pour le cas le plus complexe : celui du traitement en « **matrice ordonnable** »($i \in \{1,2\}$)

(R1) <niveau 1 i> ::= <choix niveau 1 i> **TQ(** ¬(arrière))* <action selon choix niveau œ i> <nouveau choix niveau 1 i> **FTQ***

(R2) <choix niveau 1 i> ::= **insérer|glisser|nommer|masquer|découvrir|** <modifier les choix de codage visuel i> **|autre opération|arrière|débloquer**

(R3) <action selon choix niveau 1 i> ::= **dissymétriser*** <indiquer éventuellement une composante> <blocage éventuel> <action A_i> **|non*** <action B_i>

(R4) <nouveau choix niveau 1 i> ::= **continuer|changer** <choix niveau 1 i>

(R5) <indiquer éventuellement une composante> ::= **continuation ou blocage*|ni continuation ni blocage*** <choix de composante>

(R6) <blocage éventuel> ::= **blocage ou continuation*|ni blocage ni continuation*** <choix de blocage>

(R7) <choix de blocage> ::= **bloquer|ne pas bloquer**

(R8) <action A_i> ::= **(insérer i*|** <glisser i*> | <nommer i*> | <masquer i*> | <découvrir i*>

(R9) <action B_i> ::= <modifier les choix de codage visuel i*> | <niveau 2 i*>

ETC.

C6 - Le niveau 1 de la partie dynamique

- C'est le niveau des **opérations concernant une partie**, parfois élémentaire, **du graphique**, opérations **très utiles** (voir VI.3.).
- L'opération « **nommer** » sert à connaître le nom d'un objet ou d'un caractère. Comme elle doit n'être utilisée qu'**exceptionnellement** (le traitement est visuel), elle peut éventuellement être placée au niveau 2.
- L'opération d'**insertion** est l'**opération fondamentale** : on ne peut s'en passer, elle peut suffire. Les différents reclassements automatiques peuvent être « traduits » en une suite d'insertions. L'opération de « **glissement** » est un cas particulier d'insertion. L'insertion sert à **rapprocher** les lignes et/ou les colonnes qui **se ressemblent visuellement** (voir Doc. 22).
- L'opération de « **masquage** » « **découverte** » permet à l'utilisateur de **se concentrer sur une partie du graphique** (voir Doc. 23) : -sans être gêné par le « bruit » occasionné par les données manquantes (ex.1) -pour analyser des objets selon une partie des caractères (ex.2) -pour mettre en évidence plusieurs « systèmes » de données (objets et caractères) (ex.3) - etc.
- **Modification du codage visuel** (voir X.) : par exemple : -inverser noir et blanc pour une ou plusieurs lignes ou colonnes si les données sont binaires - modifier la fonction de « paliérisation » d'une ligne ou colonne, ou de plusieurs d'entre elles si des paliers sont définis- ; il s'agit de modifications **locales, partielles ; un accès aux données** s'impose : il peut être **partiel**.
- **« Symétrie » de certaines opérations : Idée** = la plupart des manipulations envisagées sont parfaitement « **symétriques** » **(en termes de « lignes » et « colonnes »)** dans leur « **définition** » mais « **dissymétriques** » dans leur « **exécution** » : explicitons :
  - l'insertion d'une ligne est parfaitement « symétrique » de celle d'une colonne, de même le masquage, etc. ; on ne voit pas pourquoi écrire des algorithmes différents pour les lignes et colonnes (cf. la notion de « **composante** » incluant « objets » et « caractères » définie par Bertin : [BER77], p.19) ;
  - par contre, on ne va pas choisir aléatoirement quelle composante (ligne ou colonne ?) manipuler ; un choix s'impose.

Conséquences : pour de telles manipulations (sélectionnées par le système : cf. « dissymétriser »/(R3)), il faudra avant « exécution » **déterminer la composante concernée** (explicitement ou non : voir plus bas). Les **programmes** (formels) des manipulations seront **identiques,** la **différence** viendra des paramètres réels spécifiant l'« **angle** » sous lequel le graphique est considéré (ligne ou colonne ?).

- **Choix de la composante à considérer** (quand cela s'impose) : il sera effectué avant chaque manipulation concernée **sauf si** l'utilisateur a décidé de « **continuer** » **le même type de manipulation** (par exemple une suite d'insertions) ou de « **bloquer** » **le choix d'une composante** (bloquer les colonnes pour agir provisoirement sur les seules lignes).

Pourquoi ces limitations aux choix de composante ?

- bien souvent **l'utilisateur se concentrera provisoirement sur une composante** avant de retourner à l'autre, etc. ; en « bloquant » le choix de composante, il pourra, sans interrogations répétées...et donc pénibles, sans risque d'erreur, se concentrer sur la composante libre : par exemple, passer d'« insérer une ligne » à la faire « glisser » sans avoir à rappeler sans cesse qu'il s'agit de « lignes » et non de « colonnes ».
- bien souvent **la même manipulation est répétée à plusieurs reprises sur la même composante** ; c'est particulièrement vrai pour l'insertion, la principale, la plus fréquente ; il importe donc d'éviter à l'utilisateur de répéter sans cesse son choix constant, d'où la possibilité qu'il dise simplement « continuer » ; par exemple, en fonction des proximités visuelles, il va insérer une ligne, puis une autre, puis un groupe, etc. : il dira simplement de « continuer » et n'aura qu'à désigner la(les) ligne(s) à insérer et le lien d'insertion.

G7 - Manipulations élémentaires

(R1) $\forall i \in \{1,...,11\}$:
<insérer i> ::= <définir groupe connexe i> <choisir un élément de composante i>
(R'1) <définir groupe connexe i> ::= <choisir un élément de composante i> <choisir un élément de composante i>
(R2) $\forall i \in \{1,...,4\} \cup \{7,11\}$ et $\forall X_i \in \{$<masquer i>, <découvrir i>$\}$ $<X_i>$ ::= <définir groupe connexe i>
(R3) Si $i \in \{1,7\}$ (modification locale des choix de codages visuels/données binaires): <inverser i> ::= <définir groupe connexe i>
(R4) $\forall i \in \{1,...,4\} \cup \{7,8\} \cup \{10,11\}$: <glisser i> ::= <définir groupe connexe i> **REPETER** <sens de glissement> <suite à donner> **JUSQU'A(arrêt)**
(R5) <sens du glissement> ::= **avant|arrière**
(R6) <suite à donner> ::= **arrêt|poursuite**
ETC.

C7 - Manipulations élémentaires

- Pour la définition des manipulations d'**insertion** et de **glissement** : voir Document 22.
- Pour l'**exécution** de ces manipulations (cf. VIII.3.): voir Document 25.
- Remarquer l'**analogie** d'une manipulation à l'autre : toujours définir l'(les) élément(s) à considérer.
- Remarquer la **spécificité du glissement** : le glissement est un complément de la manipulation d'insertion ; il sert à **affiner** les résultats obtenus après une série d'insertion ; l'(les) élément(s) pris en compte « glisse(nt) » sur le graphique : on essaie de voir où il(s) sera(ont) le(s) mieux placé(s) ; d'où une possibilité de **tâtonnements successifs** permise par l'exécution répétitive de l'opération (cf. (R4)).
- **oix d'un élément de composante**(cf. par exemple :(R1),(R'1))) :
  - sa mise en œuvre dépendra du **matériel disponible** (a-t-on un photostyle?) et du **type de traitement** (ex : écartement régulier des éléments si représentation en matrice ordonnable, variable sinon) ;
  - on devra cependant **imposer** que le **choix** soit **visuel** et ne se réfère jamais aux numéros des composantes...encore moins à leurs noms ;
  - il est souhaitable de prévoir un **dispositif de confirmation** des choix effectués ; un tel dispositif est envisageable à d'autres moments, tout particulièrement (cf. G5) pour les transformations globales du graphique : il vaut mieux bien réfléchir à ce qu'on fait surtout si le matériel disponible ne permet pas l'« instantanéité » des opérations demandées.

COPIE NUMERIQUE

exemple de la matrice redonnable

DOCUMENT 24

DONNEES

0	8	6.1	3
3.1	4	1	9
2	10	3	9
4	8.3	12	7

composante de publication : ligne
nbre de niveaux visuels = 4
codage standard (cf. p. 21)

COPIE CODEE

0	3	3	1
1	1	0	3
0	3	0	3
0	2	3	1

+

organisation standard (cf. C4)

1	F
2	F
3	F
4	F

ORDRE LIGNE — MASQUE LIGNE

1	2	3	4	ORDRE COLONNE
F	F	F	F	MASQUE COLONNE

ORGANISATION

COPIE NUMERIQUE

insertion

seule l'ORGANISATION est modifiée

1
4
2
3

ORDRE LIGNE

etc ...

modification du codage visuel
(nbre de niveaux visuels → 2)

seule la COPIE CODEE est modifiée

0	1	1	0
0	0	0	1
0	1	0	1
0	1	1	0

etc.

masquage

seule l'ORGANISATION est modifiée

F	V	V	F

MASQUE COLONNE

etc.

ETC.

VIII. L'EXECUTION DES CHOIX DE L'UTILISATEUR

1. Introduction

Elle se décompose en la **hiérarchie** :
exécution numérique ⟶ exécution graphique (voir VI.4).

2. L'exécution numérique

Nous ne nous intéresserons ici qu'aux choix ayant un prolongement graphique.

1. Phase statique :

En fonction du traitement choisi et des choix de codage visuel effectués, le système calcule la « **copie codée** » des données numériques ; cette copie est calquée sur l'organisation qu'a le tableau en mémoire de traitement.

En fonction des choix d'organisation effectués par l'utilisateur (cf. G4), le système définit **l'organisation** initiale du graphique.

Il dispose alors de la **copie numérique** du graphique, sachant que : **copie numérique = copie codée + organisation.**

Le niveau « exécution numérique » peut alors appeler le niveau « exécution graphique » qui « traduira » la copie numérique (initiale) en graphique (initial).

Le Document 24 présente ces notions à travers un exemple de matrice ordonnable ; les variables nécessaires dans les autres cas peuvent différer, le principe général reste le même.

2. Phase dynamique :

a) Manipulation touchant à l'organisation : insérer, glisser, masquer, reclasser ; le niveau « exécution numérique » **met à jour**, en fonction de la manipulation, **la(les variable(s) décrivant l'organisation** (voir Document 24) .

b) Manipulation touchant au codage visuel : inverser noir et blanc, changer la fonction de « paliérisation », le nombre de niveaux visuels ; le système **met à jour** alors, en fonction des choix effectués la copie codée du graphique (voir Document 24). [De plus, si une variable contenant les « fonctions » décrivant les choix visuels existe, cette variable est mise à jour (voir X.4.)].

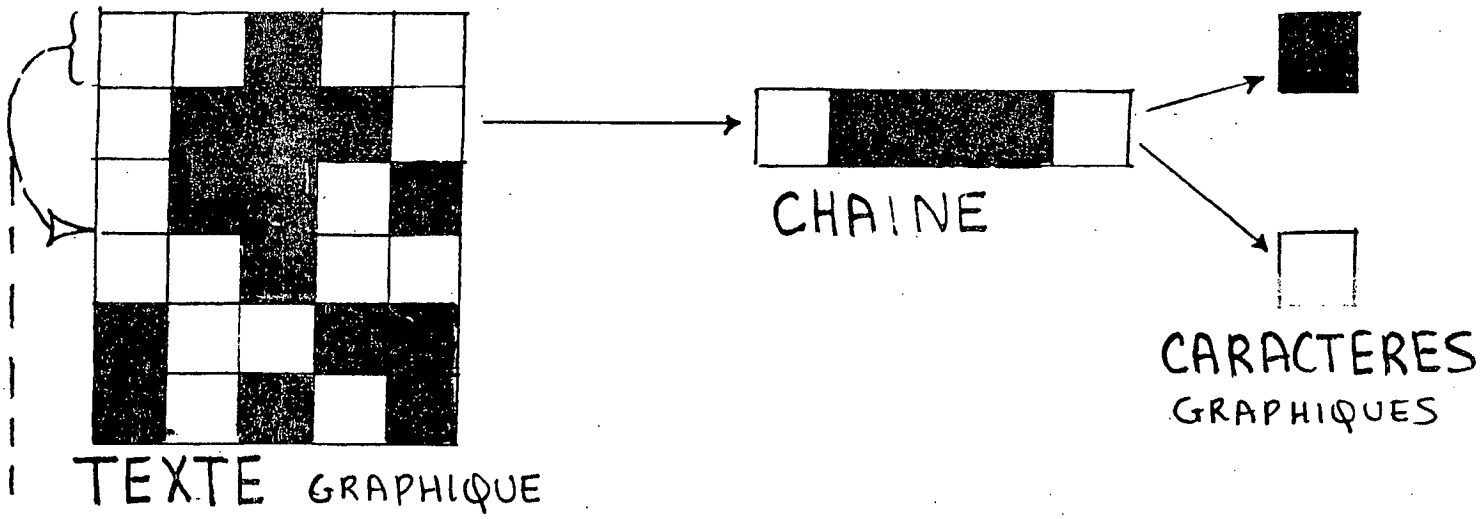
c) La copie numérique ayant été mise à jour soit sous son côté « copie codée » soit son côté « organisation », **le niveau « exécution graphique » est appelé** pour exécuter graphiquement cette mise à jour i.e. modifier le graphique en conséquence.

3. L'exécution graphique

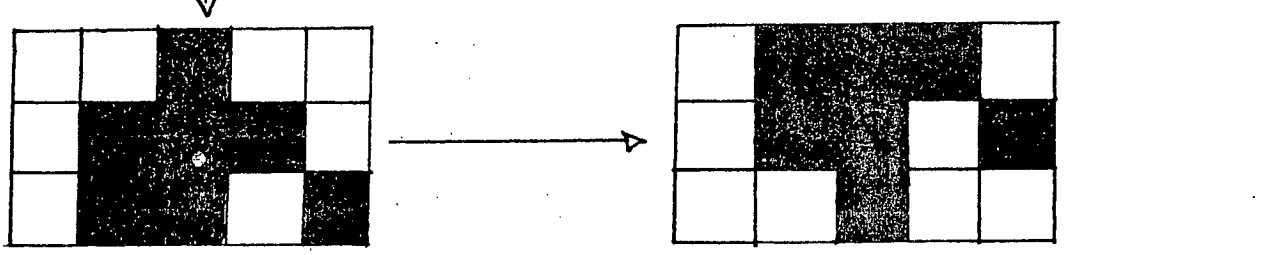
1. Deux optiques : le niveau « exécution numérique » ayant averti le niveau « exécution graphique » de : -quelle partie modifier ? -comment la modifier ?-, ce dernier peut procéder de deux façons : par **regénération** de la partie modifiée, parfois par déplacement de « fenêtres » d'écran.

2. Le choix de l'optique « regénération de la partie à modifier »: pourquoi ?
- **homogénéité** : solution possible quelque soit la manipulation ;
- si représentation en **matrice ordonnable** : on peut assimiler le traitement à un **traitement de texte** (voir infra) ;
- si représentation en **matrice pondérée** ou en **éventail de courbes**, les dessins (au trait) à regénérer sont **assez simples** ; la regénération peut **donc** être **rapide**.

MATRICE ORDONNABLE ≡ TRAITEMENT DE TEXTE



INSERTION ≡ CORRECTION



Notes : • l'insertion est prise à titre d'exemple (autres manipulations possibles)

• les données auraient pu être non linéaires ...

• l'insertion de colonnes ne pose pas de problème : les chaînes corrigées sont seulement plus courtes !

3. Assimiler le traitement en matrice ordonnable à un traitement de texte :

On considère que le graphique global ou la partie à modifier de ce graphique est un **paragraphe fait de chaînes de caractères graphiques** (voir document 23). La génération du graphique initial revient à l'**impression** du texte à l'écran, une modification graphique à une **correction** de ce texte...pouvant aller jusqu'à sa ré-impression globale si la modification est globale.

Il est clair que ces corrections par regénération des caractères utiles ne seront quasiment instantanées que si l'on a un **dispositif câblé de génération de caractères graphiques** (voire de chaînes...) ; sinon il faut simuler cette génération.

Mais il est également clair que le traitement en **matrice ordonnable** sera implantable même sur un **système ne comprenant pas de réelles possibilités graphiques** ; il suffira qu'il contienne un **système de traitement de texte** et un **jeu de caractères graphiques** (8 : c'est largement suffisant !)!.

Or le traitement en matrice ordonnable est le traitement matriciel de base (cf. IV.4.) ; on peut donc envisager de développer des **versions simplifiées du logiciel** contenant tous les traitements à représentation en matrice ordonnable, sur des systèmes sans possibilités graphiques ; ces versions seraient tout à fait convenables.

Il est clair enfin qu'en développant de telles versions, on peut **faciliter l'insertion du logiciel dans des ensembles « bureautiques » en vue de la création de « business graphics »** (voir XI.2.). Cela n'empêche pas de souhaiter que ces ensembles bureautiques contiennent des possibilités graphiques convenables !.

4. Matrice pondérée, éventail de courbes : pour générer ou regénérer de tels graphiques, il faut un minimum de primitives graphiques : celles qui permettent le **dessin au trait**.

IX. L'ASPECT « DEVELOPPEMENT »

Cet aspect doit toujours être en vue lors d'une étude de « faisabilité » comme celle-ci.

1. Ce qu'implique un développement

1. Une phase d'expérimentation : d'abord par des « spécialistes » des traitements Bertin, ensuite (ou parallèlement) par des utilisateurs quelconques ;

2. des corrections à tout moment ;

3. la possibilité d'implanter le logiciel sur des matériels ariés : capacité-mémoire, vitesse d'exécution, primitives graphiques disponibles, type d'imprimante, définition d'écran, clavier à menu, photostyle, etc ;

4. la possibilité d'offrir une gamme de versions plus ou moins évoluées : en fonction des besoins, des moyens.

2. Ce qui peut le faciliter

Ce sont les conséquences au niveau de la structure du logiciel de l'analyse du problème :

- organisations hiérarchiques imbriquées, séparation nette des différents aspects avec spécifications claires (cela permet une écriture séparée des différentes parties) .
- description formelle des choix de l'utilisateur à l'aide d'une grammaire (facilité de lecture et de modification...génération automatique de compilateurs?) ;

- isolement du niveau « exécution graphique » et mise en évidence de primitives (adaptation à différents matériels, câblage) ;
- considérer un traitement particulier comme paramètre d'une « procédure » globale (le logiciel de traitement graphique) indépendante des traitements particuliers + concevoir le logiciel comme un « jeu » de procédures minuscules (conception de versions variées, adaptation à la mémoire disponible) ;
- mettre en évidence le traitement de base : la matrice ordonnable et la manipulation de base : l'insertion (définition d'une version de base) ;
- etc. : de nombreuses remarques relatives au développement émaillent le corps du rapport.

3. Un matériel adapté

1. au niveau graphique : un dispositif câblé(*) de générations de caractères est essentiel ; y ajouter la possibilité de dessin au trait, éventuellement celle de non-affichage sélectif et contrôlé d'une partie de la mémoire-écran (si on opte pour des ommandes de masquage au niveau global) ;

(*) dans l'ensemble du logiciel, en termes d'**efficacité**, le seul point sensible est la génération des caractères graphiques (en excluant les reclassements automatiques sur les tableaux les plus grands : non essentiel).

2. au niveau du dialogue : un photostyle pour désigner des parties du graphique est souhaitable ; la possibilité de superposer modes graphique et non graphique est très intéressante ; la redéfinition (d'une partie) du clavier peut s'avérer utile ; disposer d'un petit écran supplémentaire pour le contrôle serait idéal, il servirait d'aide-mémoire : les messages à affichage assez durables pourraient souvent y être dirigés, certaines opérations pourraient y être effectuées (dans le cadre par exemple de la modification dynamique des codages visuels), un « mode d'emploi » détaillé du logiciel, des traitements Bertin, des stratégies de regroupement y seraient accessibles ; une telle possibilité n'est pas abenante dans le cadre d'un ensemble bureautique.

4. L'écriture du logiciel

Il serait souhaitable de pouvoir écrire le logiciel (versions variées) :

- dans un langage du type de celui présenté ici, pour ce qui est de l'organisation des choix (dialogue) ;
- pour l'« exécution numérique », dans un langage doté de primitives matricielles ;
- pour l'« exécution graphique », dans un langage doté e primitives évoquées précédemment (caractère, chaîne, paragraphe...).

X. LES PROLONGEMENTS A ENVISAGER

1. L'état des réalisations

La structure globale du logiciel est disponible ainsi que l'essentiel des traitements matriciels (logiciel = logiciel de traitement graphique) ; le module d'aide au choix du traitement (graphique) est réalisé.

Le travail a été effectué sur un micro-ordinateur hewlett-Packard 9845B doté de possibilités graphiques (la génération de caractères graphiques a cependant dû être simulée) ; cette première version du logiciel a été écrite en Basic.

2. Les prolongements

- **une réflexion fondamentale sur le choix des paliers** quand les données ne sont pas binaires et que le traitement l'exige est **indispensable** ; elle n'est pas spécifique des traitements Bertin : les mêmes questions se posent par exemple en cartographie. oir X.4.

- une réflexion sur les **possibilités de reclassement automatique** à offrir à l'utilisateur doit être menée ; il en existe déjà (reclassement de la matrice pondérée selon un profil-ligne par exemple) ; on peut en envisager dans d'autres traitements, par exemple en matrice ordonnable : ils ne seraient là que pour décharger l'utilisateur de tâches fastidieuses et de reclassements « évidents » non pour le remplacer ; il serait donc souhaitable de ne les rendre accessibles qu'au niveau 2. Dans ce domaine, comme d'ailleurs dans l'ensemble du logiciel, la difficulté ne réside pas dans l'écriture des algorithmes élémentaires, mais dans l'insertion judicieuse des commandes.
- l'informatisation des **traitements graphiques classiques** « statiques » (diagrammes,...) semble aisée et être rapidement réalisable ; celle des traitements graphiques non matriciels semble poser par contre des problèmes spécifiques.
- **vers l'amont dans le logiciel global** : définir des **opérateurs** permettant de **transformer globalement le tableau de données** (transposition, passer en pourcentages, centrer les variables...) ou de **calculer automatiquement des indicateurs**...en se servant par exemple des « fonctions » de codage visuel définies dans un traitement précédent (cf.: le « retour en arrière »: III.6.). Ainsi serait assurée la **connexion avec** le travail déjà réalisé à l'IMI de Nantes pour **l'aide à la constitution du tableau à traiter**.
- vers l'aval : réfléchir sur le « **graphique de communication** », réaliser le sous-module d'aide à la communication de résultats issus de traitements Bertin, le **connecter** au sous-module déjà réalisé **d'aide à la cartographie**. Voir III. et X.3.

On disposerait alors (modulo la saisie des données) non pas encore du logiciel global d'aide au traitement de données, mais déjà par contre **d'une chaîne complète** (du « monde extérieur » au « monde extérieur ») **permettant la réalisation de graphiques corrects, lisibles et synthétiques**.

3. La communication des résultats

Bertin établit une distinction très nette entre « graphique de traitement » et « **graphique de communication** » (cf. notamment [BER77] p.22) trop souvent confondus. Le premier doit transcrire exhaustivement les données et permettre le traitement ; le second vise à « la perception rapide et éventuellement la mémorisation des résultats. **Son impératif : la simplicité** ». Voir le Document 26.

En s'inspirant des idées de Bertin, on pourrait aider à la réalisation du graphique de communication (après traitement graphique matriciel) en proposant à l'utilisateur des commandes :

- **de mise en page** : séparer des blocs d'objets et/ou de caractères par des **marges** notamment ;
- **de « filtrage »** du graphique issu du traitement (représentation graphique en matrice ordonnable) : éliminer le blanc (resp. le noir) là où il semble s'apparenter à un « bruit » au milieu du noir (resp. du blanc),...;
- **d'amélioration du dessin** : estomper les effets d'escalier dans la présentation en éventail de courbes en épaississant le trait ; colorer en grisé la partie blanche des surfaces représentant des données de poids supérieur au poids moyen de leur ligne (matrice pondérée) ;
- voire de **réécriture schématique globale du graphique** : définir quelques blocs polygonaux, leur associer un certain niveau de gris, ou une certaine texture.

On rejoint alors les commandes nécessaires à la préparation de la **cartographie automatique des résultats** quand ceux-ci ont un sens topographique (voir Document 26). :

- définir des groupes d'objets ou de caractères,
- leur associer un « grain » ([BER77] p.215).

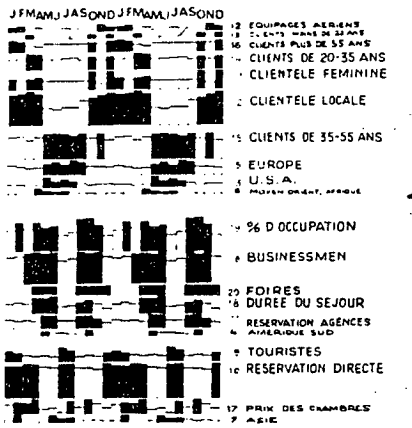
Noter que l'aide à la communication des résultats est fournie par un module séparé à partir duquel on peut accéder aux résultats d'un traitement graphique stocké en mémoire de traitement.

Ce module pourra intégrer une aide à la communication de résultats obtenus de traitements multidimensionnels classiques. Cependant **la communication des résultats** issus d'un traitement « Bertin » (tout comme l'interprétation) sera singulièrement facilitée **par la nature même du traitement** : l'information « externe » (l'interprétation) est présente tout au long du traitement grâce au rôle actif de l'utilisateur.

GRAPHIQUE DE COMMUNICATION

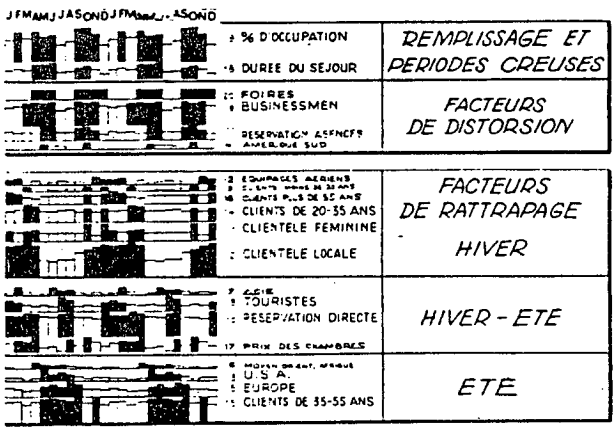
DOCUMENT 26

EXEMPLE 1



graphique de traitement

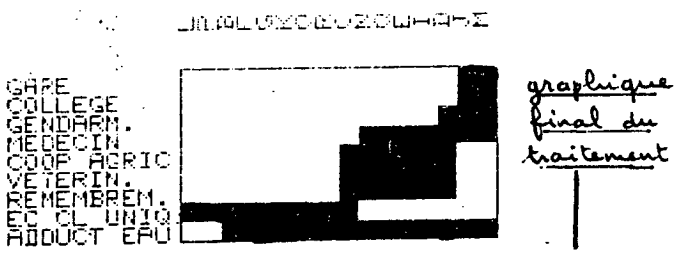
(état final)



graphique de communication

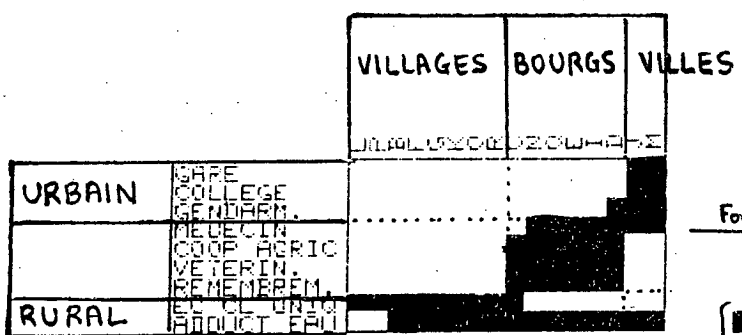
(extrait de [BER 77] p 8)

EXEMPLE 2



graphique final du traitement

[exemple voisin de l'ex.1 du <DOC237>]

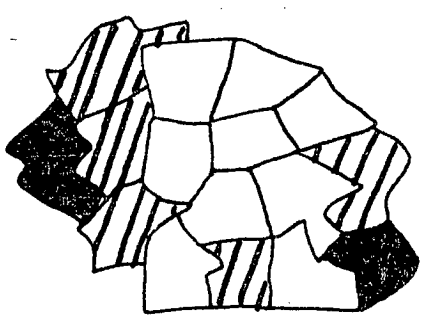


Services et équipements

GRAPHIQUE DE COMMUNICATION

Fond de carte + grain :

- VILLES
- BOURGS
- VILLAGES



CARTE DE COMMUNICATION

4. Les choix de codage visuel : définition des paliers

* Le problème de définition des paliers est **central en analyse de données** ; les résultats obtenus d'un traitement peuvent largement dépendre du codage adopté.

* Ici, la définition de paliers s'impose pour toute représentation en matrice ordonnable dès ue les données ne sont pas binaires (en simplifiant).

Le Document 21 présente une solution « standard ».

Une fois une composante de « paliérisation » choisie (cf. Document 21), **les choix de codage visuel dépendent :**

- du nombre de niveaux visuels distincts envisagés (une limite supérieure à ce nombre est fixée par la taille des matrices élémentaires, mais elle n'est pas gênante) : voir Document 21 ;
- **des « fonctions » de « paliérisation »** choisies pour chaque élément (une ligne ou une colonne) de la composante ; soit C_i un élément de composante d'indice i, $\min(C_i)$ et $\max(C_i)$ les extréma sur cet élément, une fonction de « paliérisation » fi sera définie ainsi :

fi : $[\min(C_i), \max(C_i)] \rightarrow \{1,2,\ldots,\text{Nombre de niveaux visuels}\}$.

Or, de telles fonctions peuvent être très diverses, les codages seront donc variés, les transcriptions graphiques aussi (voir Document 27).

* **Les problèmes posés sont alors les suivants :**
  - élément par élément, **quelle fonction choisir ? selon quels critères ?** notons que ces fonctions doivent conserver l'ordre des données (éventuellement l'inverser) : c'est à la base des traitements Bertin ;
  - **comment aider l'utilisateur au choix ?** comment organiser ses choix ? quand les permettre ?

Il s'agit de définir les fonctions les plus utiles, les plus élémentaires, éventuellement de prévoir leur combinaison ; d'aider l'utilisateur en lui proposant des solutions visuelles et interactives : affichage du diagramme de répartition de la ligne ou colonne considérée, possibilité de tester plusieurs fonctions et de comparer leurs conséquences en termes de codage visuel (affichage de la ligne ou colonne de matrice ordonnable concernée en plusieurs versions dépendant de la fonction choisie), possibilité de définir « manuellement » les paliers, calcul d'indicateurs relatifs aux choix effectués, etc...

* **La possibilité de redéfinir dynamiquement les paliers** (en cours de traitement) **semble idéale** : ce sera le propre des **versions les plus évoluées** du logiciel. En effet, c'est là que l'utilisateur peut, sur la base d'un certain nombre de reclassements, le mieux percevoir les améliorations à effectuer dans le choix des paliers. C'est pourquoi au départ, sauf reprise d'un traitement antérieur, le système assurera un codage visuel « standard » (cf.Document 21), voir VII.G4/C4). La redéfinition dynamique des paliers pourra toucher un élément de composante (une ligne, une colonne), un groupe d'éléments (connexes ou non), la globalité du graphique. Le choix d'une fonction pourra être commun à plusieurs éléments. La redéfinition dynamique des paliers, c'est aussi la possibilité en cours de traitement de modifier globalement le nombre de niveaux visuels disponibles, faisant ainsi varier le **contraste** de l'image (voir Document 27).

* Les choix de paliers (fonctions) doivent être mémorisés par le système pour être stockés le cas échéant. Il importe donc de pouvoir **décrire formellement ces choix** (gain de place-mémoire, absence d'ambiguité, réutilisation). De plus, une description formelle, « fonctionnelle », de ces choix, permettra de les réutiliser lors d'un retour en arrière dans la chaîne de traitement vers le niveau « aide à la constitution du tableau de données ». **Les fonctions définies pourront être reprises directement pour calculer automatiquement de nouveaux indicateurs**. On peut aussi noter que l'aide à la définition d'indicateurs pourra s'appuyer sur des commandes et un dialogue voisins de ce qui est nécessaire pour la redéfinition dynamique des paliers.

[**Remarque** : dans l'état actuel de réalisation, il est possible de modifier dynamiquement les choix de codage visuel en ne prenant plus en compte (ou en reprenant en compte) des valeurs extrémales « parasites » (cf.Document 27)].

LE CHOIX DES PALIERS

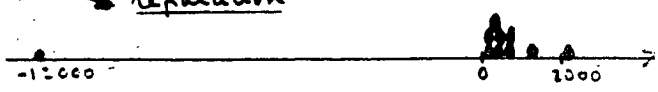
CHOIX DE LA FONCTION

(5 niveaux visuels)

DONNEES

-12000	100	500	700	1102	125	178	285	412	2100

→ répartition



codage "standard"



codage "standard"
sans prendre en compte le minimum



codage "logarithmique"
sans prendre en compte le minimum



ETC.

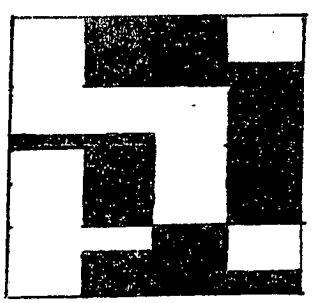
l'aspect visuel change selon la fonction de "paliérisation" choisie

un impératif : "utiliser toute la longueur de la variable visuelle" [ce n'est pas le cas de ① !]

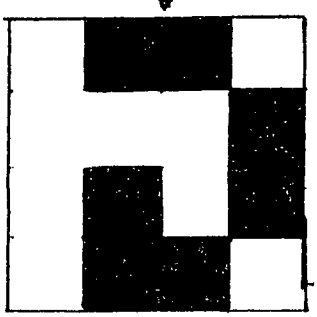
CHOIX DU NOMBRE DE NIVEAUX VISUELS

DONNEES

0	8	6.1	3
3.1	4	1	9
2	10	3	9
4	8.3	12	7



Nombre de niveaux visuels = 4



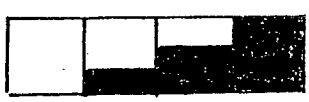
Nombre de niveaux visuels = 2

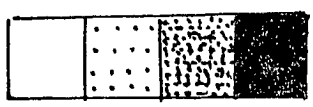
dans les deux cas : codage "standard" sur les lignes

le contraste varie

EXEMPLES DE PALETTES DE CARACTERES GRAPHIQUES

(ici : 4 niveaux visuels)





il est inutile de trop augmenter le nombre de niveaux disponibles sinon ils ne sont plus discriminants

XI. CADRE D'UTILISATION DU LOGICIEL

1. Une utilisation « grand public »

Un tel **logiciel** d'aide au traitement des données (incluant donc l'aide au traitement graphique débouchant sur la production de « business graphics ») est **destiné à des non spécialistes de l'analyse de données et des mathématiques** en général. Il a une vocation « grand public » : aide au traitement, aspect visuel, grande interactivité, présentation pédagogique (notamment analogie entre les traitements),...utilisation sur micro-ordinateur.

Il doit permettre aux utilisateurs d'élargir leur « culture statistique », en général limitée aux tableaux d'au plus trois caractères de par les limites de la graphique classique « statique » (cf.[BER77] p.24 à 27).

2. Insertion dans un ensemble bureautique pour la production de « Business Graphics »

C'est un **cadre d'utilisation essentiel.** Le logiciel global peut ne pas être inséré ; ce qui importe surtout, c'est la chaîne passant par les traitements graphiques et permettant de créer des « business graphics ».

Insérée dans un ensemble bureautique, cette chaîne constituerait un **outil précieux d'« aide à la décision »** :

- création de « business graphics » convenables par les « rédacteurs » d'articles, de rapports divers ;
- création de « business graphics » convenables par les « consommateurs » d'articles, de rapports divers contenant trop souvent des présentations incorrectes de données : possibilité de reprendre l'information présente dans ces constructions incorrectes pour obtenir des constructions convenables, immédiatement parlantes.

Ce qui peut favoriser cette insertion a été régulièrement noté au cours de ce dossier.

3. Utilisation pédagogique

1. pour la découverte de l'analyse de données : grâce aux conseils fournis par le logiciel, à la « transparence » des traitements graphiques, à l'absence de pré-requis mathématiques pour son utilisation, ce pourrait être un instrument fort utile. n tant qu'acteur du traitement, **l'utilisateur s'imprègne naturellement des principes de l'analyse de données :** réaliser les regroupements pertinents des objets, des caractères.

2. pour la pédagogie d'autres domaines : grâce à son interactivité poussée, le logiciel peut être un auxiliaire précieux pour une **démarche expérimentale**, de découverte par tâtonnements successifs, bifurcations, retours en arrière... Il est utilisable dans de nombreux domaines : Bertin cite par exemple une utilisation dans l'enseignement supérieur de la chimie (découverte de la classification de Mendeleïeff, [BER77] p.53), une expérimentation (menée avec Roberto Gimeno) à l'école primaire (découverte de la cartographie et du climat en France : [BER80]).

4. Aide aux traitements statistiques

Les tableaux ne doivent pas être trop grands (jusque 100x100 ou l'équivalent) : c'est déjà un vaste domaine. Des chercheurs, tout particulièrement en sciences sociales, peuvent être intéressés.

BIBLIOGRAPHIE

• **Ouvrage de base :**

[BER77] Jacques BERTIN : **La graphique et le traitement graphique de l'information** », Flammarion, Nouvelle Bibliothèque Scientifique, 1977.

• **Un exemple d'utilisation pédagogique des traitements « Bertin »:**

[BER80] Jacques BERTIN et Roberto GIMENO : « **La leçon de cartographie à l'école élémentaire** », Association Cartographique Internationale, 6è Conférence Internationale, Tokyo 1980.

• **Publications dont sont extraits les exemples** (hors [BER77]) :

[TR81a] TEMPS REEL (22.6.81) « **Résultats de l'enquête exclusive « Temps Réel »/ « Sagha marketing » : la maintenance informatique.**

[TR81b] TEMPS REEL (2.11.81) « **Votre métier : marché de l'emploi** ».

[CANAL83] B. RAPPACHI, Y. SIRET « **Multics : ses meilleures notes ne se ramassent pas à l'APL** », LE CANAL, bulletin d'information du CIC Grenoble, n°95, mai 1983.

[ACAD81] RECTORAT DE L'ACADEMIE DE NANTES : « **Résultats comparés du baccalauréat de 1976 à 1981** ».

Récentes publications de l'IRISA

PI 200 **Etude générale d'un réseau constitué de deux stations hyperexponentielles**
Jean-Yves Le Boudec , 12 pages : Mai 1983

PI 201 **Langage de Dyck et groupe symétrique**
Yves Cochet , 13 pages : Juin 1983

PI 202 **On the observational semantics of pair parallelism**
Philippe Darondeau, Laurent Kott , 14 pages : Juin 1983

PI 203 **Les langages fonctionnels : caractéristiques, utilisation et mise en œuvre**
Daniel Le Métayer , 162 pages ; Juin 1983

PI 204 **On exact and approximate itterative methods for general queueing networks**
Raymond A. Marie, William J. Stewart , 27 pages : Juin 1983

PI 205 **On quantifier hierarchy and its paraphrase in a semantic representation of natural language sentences**
Patrick Saint-Dizier , 17 pages : Juillet 1983

PI 206 **Trois articles sur le traitement adaptatif du signal pour l'encyclopédie Pergamon sur l'automatique**
Albert Benveniste , 64 pages : Septembre 1983

PI 207 **Sur l'existence et l'unicité du réseau homogène à un réseau ferme de files d'attente à lois générales**
Raymond Marie, Gérardo Rubino , 44 pages ; Septembre 1983

PI 208 **Problèmes d'implémentation du langage Prolog en vue de la réalisation d'une machine Prolog (Rapport final ATP « Intelligence Artificielle »)**
Yves Bekkers, Bernard Canet, Olivier Ridoux, Lucien Ungaro , 63 pages ; Octobre 1983

PI 209 **La technique du suivi de contour en synthèse d'images et ses applications**
Gérard Hégron , 28 pages : Octobre 1983

PI 210 **A new characterization of infinitary rational languages**
Philippe Darondeau, Laurent Kott , 9 pages ; Octobre 1983

PI 211 **On the observational semantics of fair parallelism**
Philippe Darondeau, Laurent Kott , 80 pages ; Octobre 1983

PI 212 **Solution à forme produit d'un système linéaire**
J. Pellaumail , 36 pages ; Novembre 1983

PI 213 **Equations de Chapman-Kolmogorov et flots stationnaires pour des processus markoviens**
J.Y. Le Boudec et J. Pellaumail , 18 pages ; Novembre 1983

PI 214 **Distribution des interentrées et intersorties pour des réseaux à forme produit**
J.Y. Le Boudec , 39 pages ; Novembre 1983

PI 215 **Un outil informatique pour l'analyse graphique de données**
René Thoraval , 60 pages : Juin 1983

PI 216 **Perturbation de la décomposition spectrale d'une matrice hermitienne**
Bernard Philippe , 21 pages : Novembre 1983

PI 217 **L'IRISA vu à travers les stages effectués par ses étudiants de DEA (1ère année de thèse)**
Edition 1983-1984 , Daniel Herman : 28 pages Janvier 1984

PI 218 **Contribution de la classification automatique pour l'organisation et l'interrogation d'un corpus de « petites annonces »**
Philippe Peter , 32 pages ; Janvier 1984

PI 219 **A micro-computer implementation of an interactive functional programming system**
Wei Zi Chu , 22 pages ; Janvier 1984

PI 220 **Commande en boucle fermée de robots munis de capteurs extéroceptifs terminaux**
Bernard Espiau , 81 pages : Janvier 1984

PI 221 **Justification et validité statistique d'une échelle** [0,1] de fréquence mathématique pour une structure de proximité sur un ensemble de variables observées
I.C. Lerman , 48 pages : Janvier 1984

PI 222 **Spécification d'une machine de gestion mémoire pour les interpréteurs des langages logiques - version 1 (provisoire)**
Yves Bekkers, Bernard Canet, Ollivier Ridoux, Lucien Ungaro , 82 pages : Février 1984

PI 223 **Une approche à la validation des protocoles d'Enchère par la méthode des tests de spécification**
Christine Ecault , 36 pages : Février 1984

PI 224 **Présentation simplifiée d'une machine de gestion de mémoire pour les interpréteurs Prolog**
Yves Bekkers, Bernard Canet, Ollivier Ridoux, Lucien Ungaro , 20 pages : Février 1984

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

ISSN 0249 - 0803