CENTRE NATIONAL D'ETUDES SPATIALES

TOULOUSE LE 29 OCTOBRE 1979

NOTE TECHNIQUE N°1

Analyse du problème du pré-traitement des données

GROUPE D'ETUDES
DES PHENOMENES AÉROSPATIAUX NON-IDENTIFIE

P R É S E N T A T I 0 N

par Alain ESTERLE

Ce document constitue la première note technique du GEPAN. Comme l'ensemble des notes techniques à venir, il est destiné à faire le point sur l'avancement des travaux et des réflexions dans une direction donnée ou un aspect particulier du problème des **P**hénomènes **A**érospatiaux **N**on-identifiés.

Il s'adresse en priorité à la Direction du CNES et aux membres du Conseil Scientifique, aux membres du GEPAN et aux collaborateurs extérieurs, ainsi qu'à tous les Organismes ou Services qui sont, par leurs activités ou recherches, associés aux travaux du GEPAN. Enfin, il est accessible, moyennant une somme modique, à toute personne ou groupe de personnes qui exprimerait le désir d'en recevoir copie en s'adressant auprès des services de relations publiques du CNES.

Dans l'avenir, les Informations provenant du GEPAN prendront l'une des deux formes suivantes :

- note technique,
- note d'information.

Les notes techniques seront essentiellement consacrées aux travaux issus du GEPAN ou issue d'une collaboration avec le GEPAN. Ces notes seront vendues à un prix inférieur à celui de l'édition. Par contre, les notes d 'information seront plutôt orientées vers la présentation et la divulgation d'études, réflexions et travaux développés indépendamment du GEPAN (par exemple, à l'étranger, mais dont le GEPAN considère que la connaissance est importante pour tous ceux qui s'intéressent aux **P**hénomènes **A**érospatiaux **N**on-identifiés). Les notes d'information seront gratuites. Une première note d'information est actuellement en préparation et paraîtra après la première note technique.

Cette première note technique est destinée à faire le point sur un des volets essentiels des activités du GEPAN. Ces dernières peuvent être grossièrement divisées en trois parties :

- la collecte de données,
- l'analyse et le pré-traitement des données,
- le développement des études scientifiques.

L'enchaînement logique et chronologique de ces trois étapes nous interdit d'en négliger aucune au profit des autres, en raison de leurs liens évidents. Les recherches ne peuvent progresser qu'à partir de données correctement analysées et préparées, et ces opérations ne sont possibles que dans la mesure où la collecte a été soumise à des règles minimales de rigueur et d'homogénéité. En retour, les conclusions partielles des études spécifiques permettront de mieux orienter et raffiner les techniques et le contenu des collectes de données.

La présente note expose l'ensemble des problèmes liés à la phase intermédiaire d'analyse et de pré-traitement des données. Cette phase comporte plusieurs étapes :

- la saisie des documents,
- l'expertise des documents,
- leur codage et la construction de fichiers informatiques.

Cette phase débouche ensuite sur un type d'études importantes:

- les traitements statistiques.

Jusqu'à présent, le GEPAN a fait un certain nombre de choix pour ces différentes opérations (classification, règles de codage, etc...). Ces choix ont permis de développer le travail pendant deux ans et ont abouti à de premiers résultats d'analyses. Ces choix et ces travaux sont ici exposés. Cependant, une analyse critique de ces choix a été concurremment menée, tant à partir d'une réflexion théorique, qu'à partir des problèmes pratiques rencontrés. Cette réflexion a donné lieu à la rédaction de certains documents qui sont ici joints.

Ce document n'apporte pas de conclusion, la réflexion continue et conduira d'ici la fin de l'année, à la mise en pratique de principes de travail plus adaptés qui seront exposés aux personnes, organismes ou groupes amenés à s'associer au GEPAN.

Ceci donnera lieu, bien entendu, à une autre note technique, le jour où les améliorations auront été mises au point et expérimentées.

SOMMAIRE

PRÉSENTATION

CHAPITRE 1 - Les techniques d'expertise
(Alain ESTERLE)

CHAPITRE 2 - Règles de codage (4ème version GEPAN)
(Jacques DUVAL)

CHAPITRE 3 - Rapport d'études statistiques
(Paul LEGENDRE)

CHAPITRE 4 - Réflexion sur les principes du pre-traitement
(Manuel JIMENEZ)

CHAPITRE 5 - Etude des problèmes liés à la création d'un fichier informatique
(Jean-Pierre ROSPARS)

CHAPITRE 6 - Les méthodes de traitement de l'information appliquées aux phénomènes OVNI
(Michel MARCUS)

CONCLUSION

CHAPITRE 1

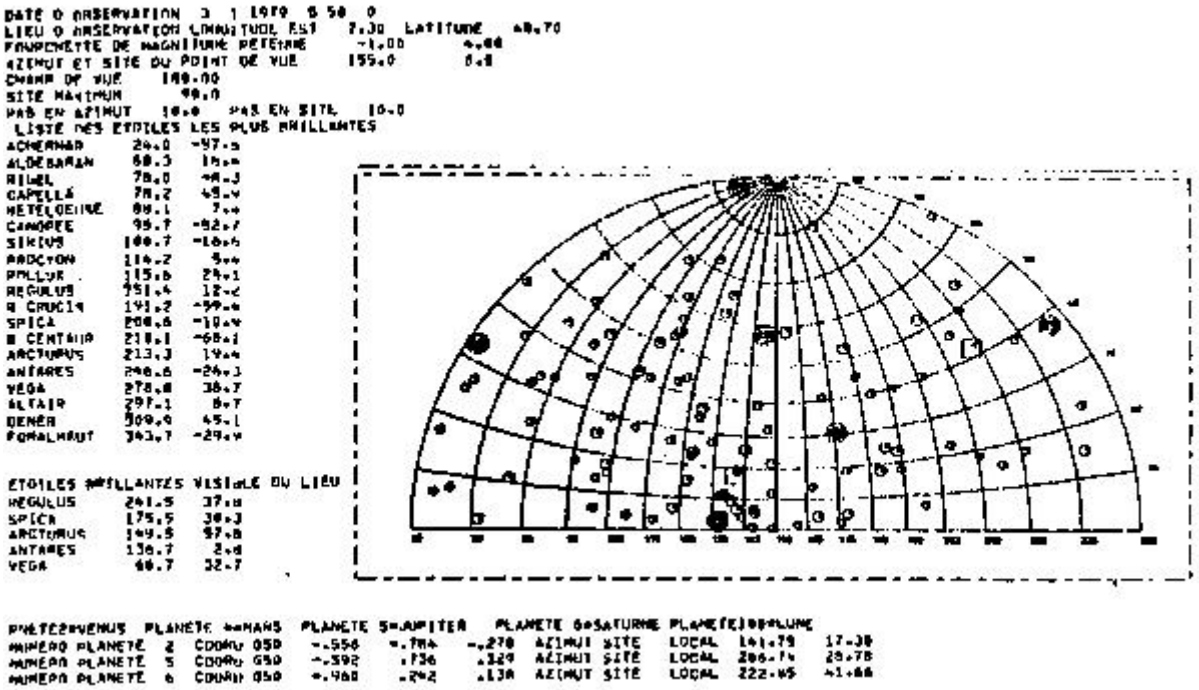
LES TECHNIQUES D'EXPERTISE

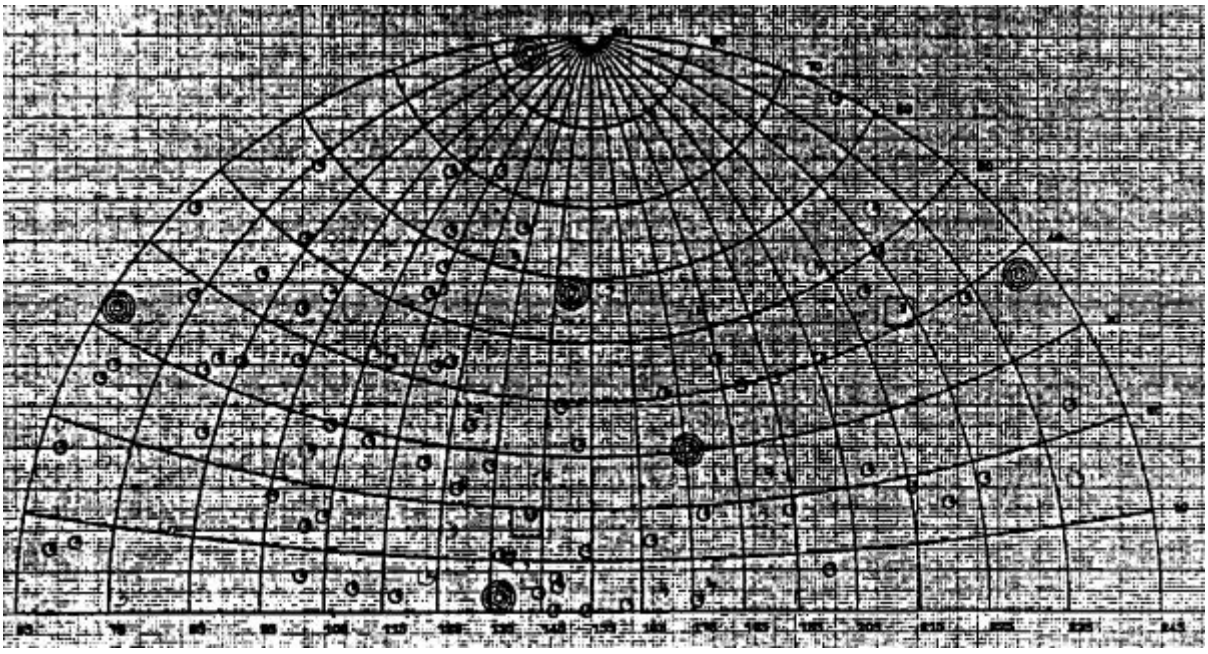
L'expertise se pratique, à l'heure actuelle, essentiellement sur les procès-verbaux de la Gendarmerie Nationale. Il s'agit là d'une source d'information spécifique qui n'avait pas auparavant été exploitée systématiquement. Les résultats obtenus, c'est à dire la distribution statistique des différentes classes (voir la plaquette "Le GEPAN et l'étude du phénomène OVNI"), sont certainement liés à la source et n'ont donc qu'une signification relative.

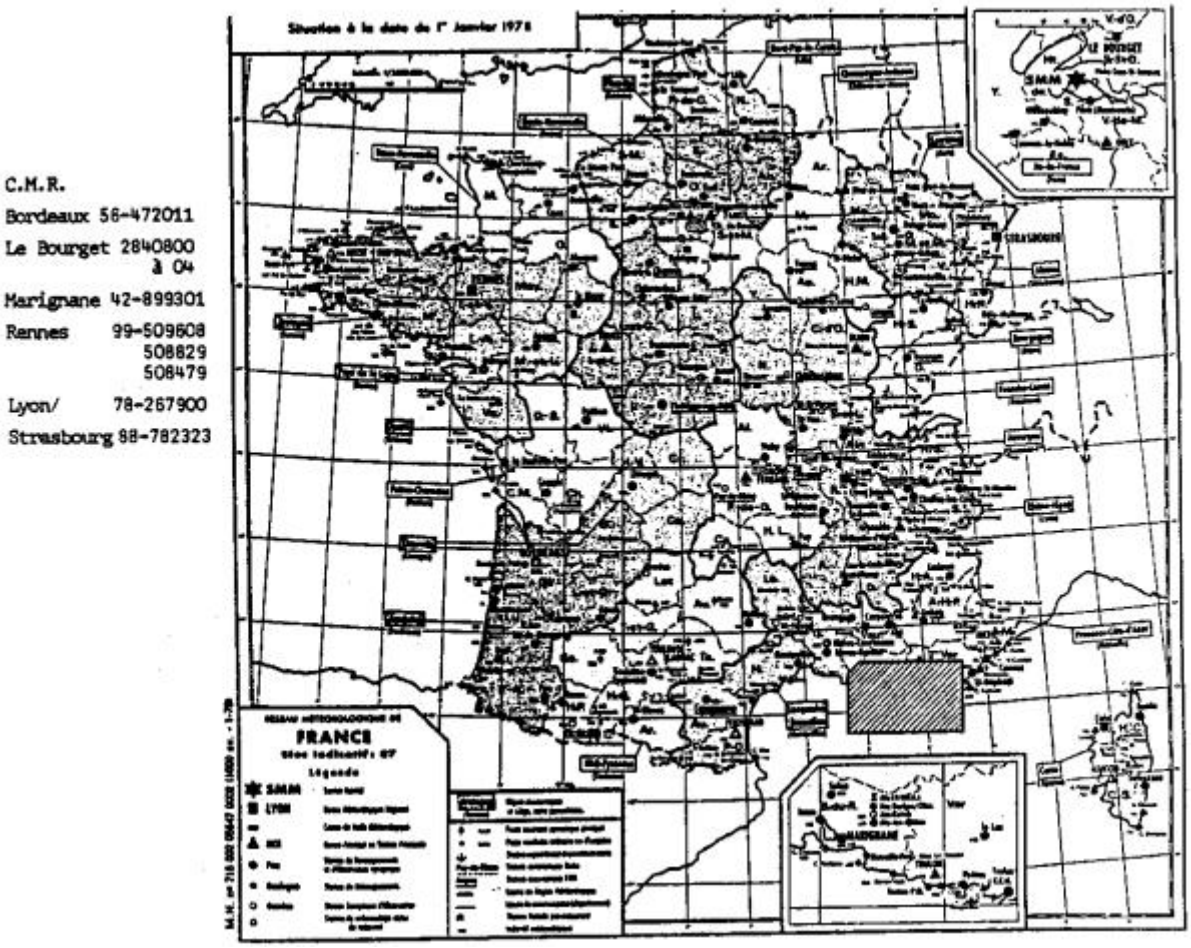
L'expertise se fait en analysant les documents de témoignage par rapport à des documents d'ordre général sur la manifestation de phénomènes connus (par exemple : les ballons, les avions, etc...). De plus, l'expertise utilise des renseignements spécifiques liés aux conditions d'observation (position des astres, données météorologiques, lignes aériennes, ...).

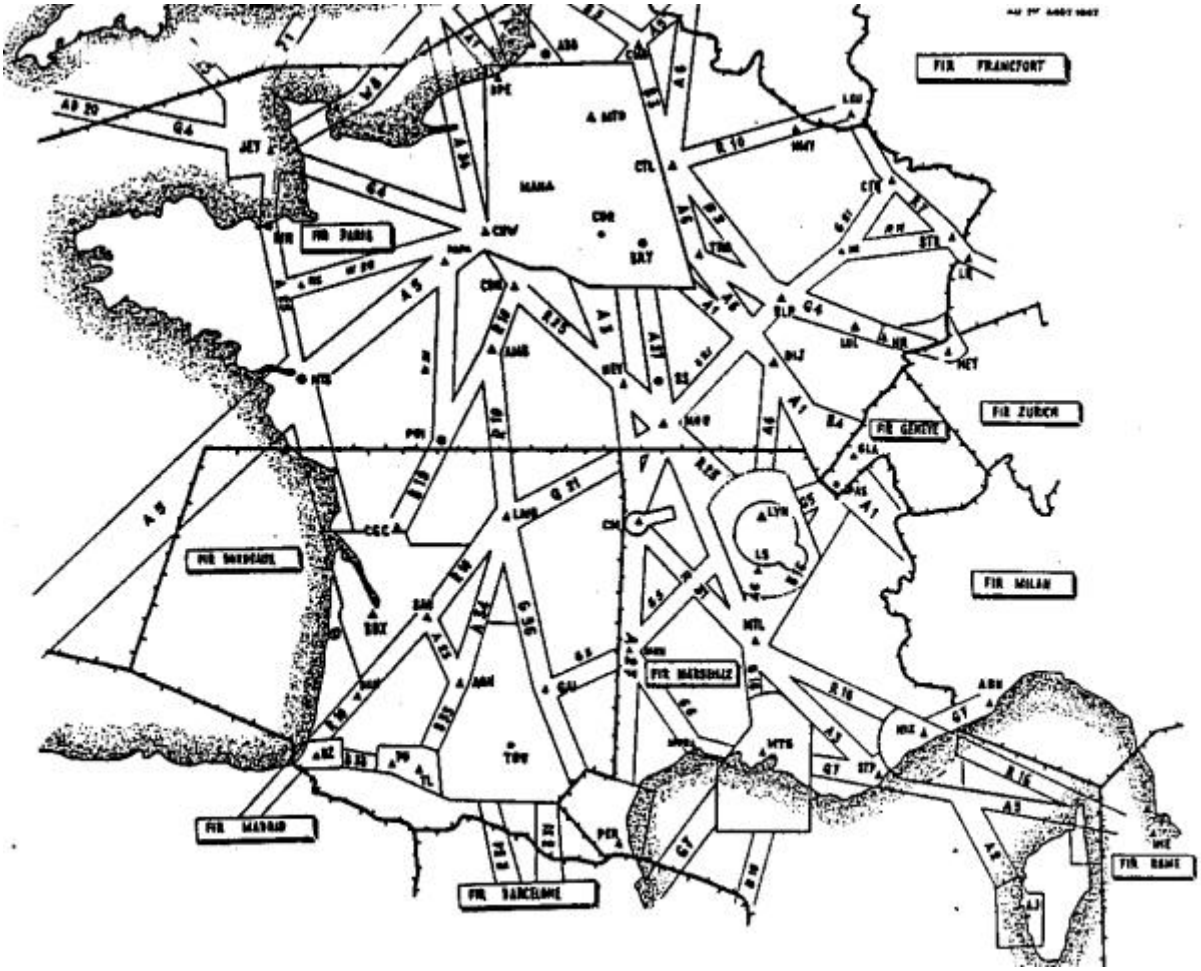
L'expertise débouche sur une classification à quatre modalités la quatrième (phénomènes non-identifiés susceptibles d'être ultérieurement étudiés) étant scindées en 6 modalités couramment appelées « Classification de HYNEK » . Sont aussi évalués les indices « d'intérêt » et de « crédibilité ».

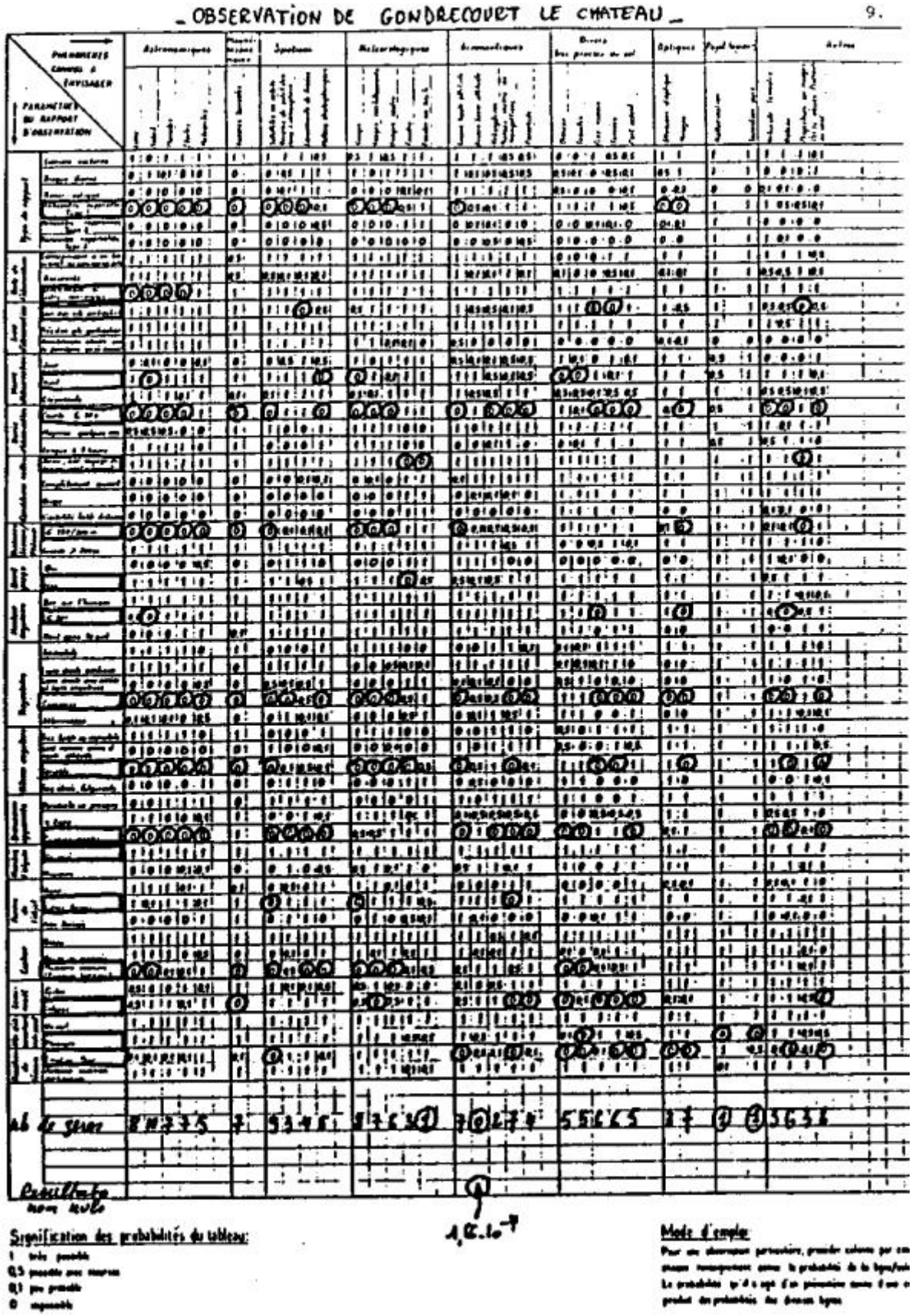
Après expertise, une « matrice de test » sert à vérifier si aucun point important n'a été oublié. Après vérification, les documents expertisés (et prioritairement, ceux de la quatrième modalité) sont codés en vue d'études statistique (voir chapitre 2). Cette technique de codage ne prétend pas résoudre le problème dans sa totalité (codage de toute l'information), mais seulement permettre d'aborder certains aspects et surtout mettre en évidence les difficultés Inhérentes à ce type de données.






C.M.R.
Bordeaux 56-472011
Le Bourget 2840800
à 04
Marignane 42-899301
Rennes 99-509608
508829
508479
Lyon/ 78-267900
Strasbourg 88-782323
FRANCE


FIR FRANCFORT
FIR PARIS
FIR ZURICH
FIR GENEVE
FIR BORDEAUX
FIR MILAN
FIR MARSEILLE
FIR MADRID
FIR BARCELONE
FIR ROME


OBSERVATION DE GONDRECOURT LE CHATEAU
9.
PARAMETRES CONNUS A EVALUER
PARAMETRES DU RAPPORT D'OBSERVATION
Astronomiques
Spatiaux
Météorologiques
Aéronautiques
Divers
Optiques
Actifs
Résultats non nuls
Signification des probabilités du tableau:
1 très possible
0,5 possible mais réserves
0,1 peu possible
0 impossible
Mode d'emploi

CHAPITRE 2

RÈGLES DE CODAGE

(4ème version)

Rédacteur : M. DUVAL

INTRODUCTION

Ce codage a pour but de permettre la constitution d'un fichier général à partir duquel des études statistiques pourront être développées.

De tels fichiers et de telles études ont déjà été mise en oeuvre dans le passé au CNES ou hors du CNES. Ces démarches ont amené de multiples réflexions sur la méthodologie statistique. Il s'agit maintenant pour nous, de profiter de ces réflexions afin de généraliser, si possible, les travaux déjà menés.

Noue ne discuterons pas ici des raisons qui nous ont conduit à choisir ce codage. Nous nous contenterons de l'expliciter en détail.

Partant de l'idée qu'un témoignage codé ne doit pas excéder le contenu d'une carte, soit 80 caractères, le codage se définit par le choix d'un jeu de critères à mettre en évidence, ainsi que l'ensemble des valeurs disponibles pour ces critères.

On en déduit le nombre de caractères à réserver pour chaque critère. Nous détaillerons d'abord la suite des critères utilisés ainsi que les caractères réservés sur la carte FORTRAN . Ensuite, nous Indiquerons le codage de chaque critère proprement dit, c'est à dire l'ensemble des signes à utiliser et la signification de chacun d'eux.

AVERTISSEMENT

Dans le cas où une observation regroupe plusieurs témoins dépendants, le choix du témoin principal ne doit pas obligatoirement être lié ni à la profession, ni à l'âge de l'un d'entre eux. La déclaration la plus détaillée, comparée à celle des autres témoins, afin d'en faire ressortir la crédibilité, peut faire retenir son auteur comme principal témoin. Il est bien entendu que ceci reste à l'appréciation du codeur.

Dans le cas où une observation regroupe plusieurs témoins indépendants, rien ne permet de dire qu'il s'agit du même objet. En conséquence, il sera bon de faire à cet effet autant de fiches qu'il y a d'observateurs (voir 2.8 : numéro du témoin).

Les rapports de gendarmerie permettent en général, de connaître les nom, âge, profession et adresse des témoins. On portera ces renseignements sur une feuille distincte des fiches de codage (FRO).

Nous décrirons dans les pages suivantes, la quatrième version du codage des cas "D". Afin de pouvoir éventuellement améliorer le codage sans autant reprendre un à un tous les dossiers d'origine, nous recommandons de remplir en clair les fiches résumé d'observation *en utilisant le plus possible les termes mêmes des rapports*, et en n'oubliant pas de cocher la case correspondante « mentionnée» ou « déduite » selon le cas.

1. CRITÈRES INTERVENANT DANS LE CODAGE

TABLEAU N°1

Critères	Nbre de caractères	N° des cases sur la carte
Numéro	8	1 à 8
Source	2	9 à 10
Type d'observation	1	11
Crédibilité	1	12
Intérêt	1	13
Code utilisé	1	14
Initiales du codeur	2	15 à 16
Numéro du témoin	1	17
Date	8	18 à 25
Heure	5	26 à 30
Département	2	31 à 32
Longitude	4	33 à 36
Latitude	4	37 à 40
Nature du lieu	1	41
Nombre de témoins	1	42
Profession du témoin	2	43 à 44
Age du témoin	3	45 à 47
Sexe	1	48
Conditions météorologiques	1	49
Durée d'observation	5	50 à 54
Distance minimale	4	55 à 58
Méthode d'observation	1	59
Nombre d'objets	2	60 à 61
Forme	1	62
Taille	2	63 à 64
Luminosité	2	65 à 66
Couleur	1	67
Trajectoire	2	68 à 69
Vitesse	4	70 à 73
Accélération	1	74
Bruit	1	75
Hauteur angulaire	2	76 à 77
Direction azimutale	2	78 à 79

2. CODE DES CRITÈRES

2.1. NUMÉRO - 8 cases

A chaque témoignage est attribué, au secrétariat du GEPAN un numéro à 8 chiffres. Pour les enquêtes de Gendarmerie, ce numéro est imprimé en grand sur la page de garde. Les deux premiers chiffres donnent l'année à partir de 1900 et le 3ème donne la source (ex. : 3=Gendarmerie nationale/métropole). On utilisera les 8 chiffres que l'on mettra dans les 8 cases prévues à cet effet.

2.2. SOURCE - 2 cases

On utilisera les abréviations suivantes :

- GN : enquêtes de la Gendarmerie nationale
- PA : pour la police de l'air
- JS : pour les journaux spécialisés
- JN : pour les journaux non spécialisés
- GP : pour les groupements privés.

2.3. TYPE DE TÉMOIGNAGE - 1 case

- A : pour les cas classés A
- B : pour les cas classés B
- C : pour les cas classés C

Pour les cas D, on utilisera les numéros correspondants suivants :

- 1 : lumière nocturne
- 2 : disque diurne
- 3 : radar-visuel
- 4 : observation rapprochée type 1
- 5 : observation rapprochée type 2
- 6 : observation rapprochée type 3

Cette qualification du témoignage fait partie de l'expertise préalable. Le résultat de l'expertise se trouve consigné dans la fiche de classification témoignage (voir plus loin). Si il y a ambiguïté, on prend le numéro le plus élevé.

2.4. CRÉDIBILITÉ - 1 case

Entre 2 et 6, on additionnera les crédibilités attribuées par les deux notateurs dans les tableaux en bas des fiches d'expertise.

- *: Pas d'indication ou incomplète
- 1 : crédibilité faible
- 2 : crédibilité moyenne
- 3 : crédibilité farte.

2.5. INTERET - 1 case

Ceci se fait comme pour la crédibilité (voir 2.4).

2.6. INDICATIF DU CODAGE - 2 cases

Les conventions de codage pourraient être ultérieurement changées et des cartes modifiées dans ce sens. Il faut donc prévoir une case indiquant quelles conventions de codage sont utilisées à un moment donné. Le code défini dans le présent document correspondra à l'indicatif 2.

2.7. INITIALES DU CODEUR - 2 cases

On donnera les initiales du prénom et du nom dans cet ordre.

2.8. NUMERO DU TEMOIN – 1 case

Plusieurs témoins, dépendants ou indépendants, peuvent avoir fait une même observation et donner des informations différentes mais intéressantes.

Dans ces cas, et ceci indépendamment du critère "nombre de témoins", numéroter les principaux, et établir une fiche pour chacun. Pour ceci, noter en clair sur la ligne le numéro attribué sur la quantité choisie et porter le numéro dans la case correspondante (ex. : 2/3 (2)).

2.9. DATE - 8 cases

Date non disponible : ********
Si la date est disponible, les deux premières cases contiennent le quantième du jour, les deux suivantes le numéro du mois et les quatre dernières, l'année (ex. 1er février 1978 = 01021978). Si la date est partiellement connue, on codera les parties connues et on mettra des * pour les parties inconnues.

2.10. HEURE - 5 cases

Utilisation des 5 caractères de la façon suivante :

a) 4 premiers : Heure locale suivant le format : hhmm
Si cette heure est inconnue : ****

b) 5ème : estimation grossière de l'heure suivant le code :
M = matin
V = vers midi
P = après midi
S = soirée
C = crépuscule
D = début de la nuit
Z = vers minuit
F = fin de la nuit
L = aurore
Exemple : 20 heures 40 minutes = 2040D
Donc, en général, l'heure sera codée selon les deux codes.

2.11. DEPARTEMENT - 2 cases

On mettra le numéro du département dans lequel se trouve le lieu de l'observation.
Exemple : Haute-Garonne = 31. Si l'information n'est pas disponible, on mettra **.

Si l'observation a été faite hors de la métropole, dans les territoires d'outre mer, on mettra deux zéros (00). Pour l'étranger, on mettra deux point (. .).

2.12. LONGITUDE - 4 cases

Il s'agit de la longitude assurée en degrés et fractions décimales (au dixième de degré près), à partir de Greenwich, positivement vers l'Est.

Exemple :
+ 2 degrés et 15 mn = +022
- 2 degrés et 25 centièmes de degré = -022 ou 3578
S'il n'y a pas d'information de longitude disponible, on mettra quatre * (****) , NOTA : en France métropolitaine, les longitudes sont comprises entre - 9 et + 9 degrés. La longitude de Paris est de 2,33 degrés.

Fiche de classification des témoignages

FICHE DE CLASSIFICATION
DES OBSERVATIONS DU TYPE "D"

CLASSIFICATION

Le phénomène non-identifié qui fait l'objet du CR n° .....
est à classer dans la rubrique suivante :

	Expert n° 1	Expert n° 2
Lumière nocturne		
Disque diurne		
Radar-visuel		
Obs. rapprochée type 1		
Obs. rapprochée type 2		
Obs. rapprochée type 3		

JUGEMENT DE VALEUR RELATIVE

Cette observation est à classer, à mon avis, dans la rubrique suivante (1), en ce qui concerne son intérêt et la crédibilité des témoins :

Haute crédibilité			
Crédibilité moyenne.....			
Faible crédibilité			

Peu d'intérêt pour une étude du phénomène
Intérêt moyen
Très grand intérêt pour une étude du phénomène

(1) L'expert n° 1 fera une croix X, l'expert n° 2 fera un cercle O dans la case correspondant au choix.

2.13. LATITUDE - 4 cases

Il s'agit de la latitude mesurée en degrés et centièmes de degré (au dixième de degré près) positivement vers le Nord. Exemple : 49 degrés et 20 minutes = + 493. S'il n'y a pas d'information de latitude, on mettra **** .

2.14. NATURE DU LIEU - 1 case

Témoins potentiels :
* ....: inconnue
0 ..... D .....: désert, haute montagne, mer
0 à 10..... S.....: habitation isolée
10 à 100..... H..... : hameau, petit village
100 à 1000.....H .....: bourgade, banlieue
1000 à 10 000..V .....: ville
10 000.....A .....: vue d'avion
.....*..... : connue mais n'entre pas dans les rubriques précédentes

2.15. NOMBRE DE TÉMOINS - 1 case

Code	*	1	2	3	4	5	6	7	8	9
Nbre témoins	inconnu	1	2	3	4	5	dizaine	centaine	millier	+ millier

2.16. PROFESSION DU TEMOIN PRINCIPAL - 2 cases

Voir annexe 2 - Si la profession est inconnue, on mettra *

2.17. AGE DU TEMOIN PRINCIPAL - 3 cases

L'age du témoin étant en général connu, il semble intéressant de le coder sur 3 caractères afin de limiter l'effet de l'arbitraire de la classification.

2 premiers caractères : âge en clair
3ème caractère : classe (suivant le code actuel suivant) :

- * : âge inconnu
- E : enfant de 0 à 13 ans
- J : adolescent de 14 à 20 ans
- A : adulte de 21 à 59 ans
- C : vieillard de 60 ans et plus.

2.18. SEXE DU TEMOIN - 1 case

On utilisera les numéros correspondants suivants :
- 1 : masculin
- 2 : féminin

2.19. CONDITIONS METEOROLOGIQUES - 1 case

- : pas d'indication
- 1 : très beau temps, ciel pur

- 2 : nuages épars
- 3 : ciel couvert mais à haute altitude
- 4 : ciel bas, mauvais temps sans pluie
- 5 : pluie, grêle, neige, orage, faible visibilité
- 6 : 1 avec vent
- 7 : 2 avec vent
- 8 : 3 avec vent
- 9 : 4 avec vent
- A : 5 avec vent
- **.** : autres conditions

2.20. DURÉE DE L'OBSERVATION - 5 cases

Ce critère sera codé en clair et de la façon suivante :

- les 4 premières cases pour la durée en chiffre
- la 5ème case pour la classe (**S** : secondes, **M** : minutes, **H** : heures).

Toutes les cases sont remplies :

- * : pas d'indication de durée
- S : de 0 à 59 secondes
- M : de 1 à 1440 minutes (soit de 1 mn à 24 h)
- H : supérieure à 24 heures
- **.** : autres cas

Exemple : 15 secondes : 0015S
2 heures 45 minutes : 0165M

2.21. DISTANCE MINIMALE D'OBSERVATION - 4 cases

Ce critère sera codé en clair et de la façon suivante :

- les 3 premières cases pour la distance chiffrée –
- la 4ème pour la classe (M-mètres, K-kilomètres, A>3 km) .

- * : pas d'indication
- M : de 0 à 999 mètres
- K : de 1 à 3 kilomètres
- A : supérieure à 3 kilomètres

Exemple : 55 mètres : 055M ; 1,5 km : 1.5K

2.22. MÉTHODE D'OBSERVATION - 1 case

- : pas d'indication
- A : oeil nu au sol
- B : Jumelles, longue vue, théodolite
- C : lunette astronomique
- D : télescope
- E : photographie ou film
- F : radar
- G : jumelles + photo
- H : visuel + radar

- J : oeil nu à partir d'un avion
- K : oeil nu à partir d'un bateau
- L : jumelles à partir d'un bateau
- M : télescope + photo
- N : à bord d'un véhicule automobile en marche
- P : à bord d'un véhicule automobile à l'arrêt
- **.** : autres méthodes

2.23. NOMBRE D'OBJETS - 2 cases

- * : pas d'indication
- 00 : aucun objet
- de 01 à 98 : nombre d'objets si inférieur à 99
- 99 : si 99 objets ou plus

NOTA : au-delà de 10 objets, on arrondit nombre de dizaines si le nombre n'est spécifié explicitement nulle part.

2.24. FORME DE L'OBJET PRINCIPAL - 1 case

- * : pas d'indication
- A : disque, soucoupe lenticulaire
- B : ronde, circulaire, boule
- C : cigare, cylindrer fusée
- D : oeuf, ovale, ovoïde, ballon de rugby
- E : conique, triangulaire, chapeau asiatique, trapézoïdale
- F : toupie
- G : carrée, rectangulaire, parallélépipédique
- F : soucoupe à coupole, chapeau de canotier
- J : couronne, pneumatique
- K : ponctuelle, étoile, grosse planète
- L : dôme, tasse, parachute, parapluie, meule de foin
- M : méduse,, champignon
- N : croix
- P : croissant
- Q : cigare accompagné de disques
- R : nuée, nuage, halo
- S : nid d'abeilles
- **.** : autres formes

2.25. TAILLE - 2 cases

1er cas : évaluation non métrique

Pour la 1ère case :

- C - par comparaison
- A - angulaire
- - pas d'indication
- **.** - autres types non métriques

Pour la 2ème case :

- A - immense, très gros

Par comparaison
Exemple :
comme la lune = CG

- H - comme une pièce de 5F
- C - comme une orange
- D - comme une assiette, un melon
- B - comme une citrouille
- F - comice un avion
- G - comme la lune
- H - comme une voiture
- J - comme une grosse étoile
- 0 à 9 dizaines de minutes d'arc
- **.** non codable (sup. à 1° 3')

Angulaire, exemple A6 (1°)

2ème cas : évaluation métrique
on utilise les 2 cases pour coder la plus grande dimension évaluée :

- 01 à 98 : taille en mètres
- 99 : 99mètres ou plus
- 00 : inférieur à 1 mètre

3ème cas : pas d'indication, on utilisera : **

NOTA : Dans le cas où plusieurs types d'indications sont fournis, on gardera l'indication la plus crédible, cette appréciation étant laissée au codeur. De toutes façons, les deux cases doivent être remplies.

2.26. LUMINOSITÉ - 2 cases

Combinaison de 2 paramètres si nécessaire :
- 1ère case : paramètre paraissant le plus important,
- 2ème case : paramètre apportant un enrichissement.

Si un seul paramètre, laisser la 2ème case vierge.
- * : pas d'indication
- 1 : lueur, faiblement lumineux
- 2 : lumineux, fluorescent
- 3 : brillant, très lumineux
- 4 : intense, éblouissant, éclatant
- 5 : non lumineux
- 6 : réfléchit la lumière du soleil on autre lumière
- 7 : halo seulement
- 8 : variable en intensité (de O à infini mais non périodique)
- 9 : clignotant
- A : non lumineux mais avec faisceaux
- **.** : autres types

2.27. COULEUR - 1 case

- * : pas d'indication
- A : rouge sombre
- B : rouge
- C : orangé, feu

- D : jaune, ambre
- E : vert
- F : bleu
- G : bleu sombre, métallique, indigo
- H : violet
- J : blanc
- K : noir
- L : gris
- M : métallique (argent, aluminium poli)
- N : plusieurs couleurs
- P : couleur (s) changeante (s)
- Q : marron
- R : or
- **.** : autres couleurs

2.28. TRAJECTOIRE - 2 cases

Combinaison de 2 paramètres si nécessaire. Si un seul paramètre, laisser la 2ème case vierge.

- * : pas d'indication
- A : ligne droite, ou courbe très ample - immobile ou ligne droite avec arrêts
- B : virages brusques
- C : arabesques compliquées
- D : trajectoire complexe mais analysable (périodicité, suivi de route, de fleuve, etc...)
- E : stationnement près du sol
- F : atterrissage et arrêt prolongé avant décollage
- G : atterrissage puis décollage immédiat
- H : objet vu au sol qui décolle
- J : objet pénétrant ou sortant de l'eau
- K : objet qui monte et se perd dans les étoiles
- L : nulle puis lente
- M : nulle puis rapide
- **.** : autres types de trajectoire

2.29. VITESSE - 4 cases

Ce critère sera codé en clair de la façon suivante :

- les 3 premières casse pour la vitesse chiffrée (en centaines Km/h)
- la 4ème case, pour la classe.

- * : pas d'indication
- A : lente ou très lente ou immobile
- B : très rapide
- C : variable
- D : fulgurante
- E : vitesse d'un avion
- **.** : autres types de vitesse

Si l'information est variable, on code la plus grande suivie de la lettre C.

Exemple :

- très rapide 1000 km/h : 010B
- immobile : 000A
- variable sans précision : ***C

2.30. ACCELERATION - 1 case

- * : pas d'indication
- 1 : faible
- 2 : variable
- 3 : élevée
- **.** : autres types

2.31. BRUIT - 1 case

- * : pas d'indication
- A : aucun bruit, silence total, objet silencieux
- B : bourdonnement, vrombissement, bruit d'abeilles, grondement
- C : sifflement aigu
- D : bruit d'air comprime s'échappant
- E : bruit de moteur électrique démarrant ou de machine centrifuge
- F : explosions violentes
- G : bruit de vent violent sous l'objet
- H : aucun bruit perceptible
- **.** : autres bruits

2.32. HAUTEUR ANGULAIRE - 2 cases

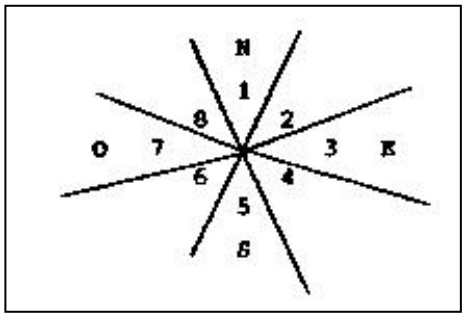
On notera la hauteur en début et en fin d'observation :
- 1ère case : début
- 2ème case : fin

Si un seul paramètre, laisser la 2ème case vierge.

- * : pas d'indication
- 1 : de 0 à 15° (ou bas sur l'horizon)
- 2 : de 15 à 30°
- 3 : de 30 à 45°
- 4 : de 45 à 60°
- 5 : de 60 à 90° (proche du zénith)
- 6 : au-dessous de l'horizon, sous un avion
- 7 : observé d'un avion, à la même hauteur ou au-dessus
- 8 : objet vu au sol ou près du sol
- **.** : autres types

2.33. DIRECTIONS AZIMUTALES - 2 cases

Les directions azimutales seront estimées dans les secteurs angulaires suivants : (azimut en degrés à partir du Nord, positivement vers l'Est) :



- * : pas d'indication
- 1 : -22.5 à + 22,5 vers le nord
- 2 : 22,5 à 67,5 nord-est
- 3 : - 67,5 à 112,5 est
- 4 : 112,5 à 157,5 sud-est
- 5 : 157,5 à 202,5 sud
- 6 : 202,5 à 247,5 sud-ouest
- 7 : 247,5 à 292,5 ouest
- 8 : 292,5 à 337,5 nord-ouest
- 0 : à la verticale (vers le zénith) -
- **.** : pas codable

On donne ainsi la direction azimutale au début de l'observation puis en fin d'observation.

3. CONCLUSION

En règle générale, quand l'information n'est pas disponible, on code ***** et quand l'information est disponible, mais non compatible, avec les règles de codage prévues, on code **.** (point). Par conséquent, il ne peut pas en principe, y avoir de blancs dans les 79 cases exceptées les cases 66, 69 et 77.

Il s'agit donc, en général, d'une grille de signification. Les indications contenues dans les témoignages peuvent n'être identiques à aucun des cas proposés, mais se rapprocher fortement de certains. Ceci reste à l'appréciation du codeur.

LISTE GROUPES ET DES CATÉGORIES SOCIOPROFESSIONNELLES

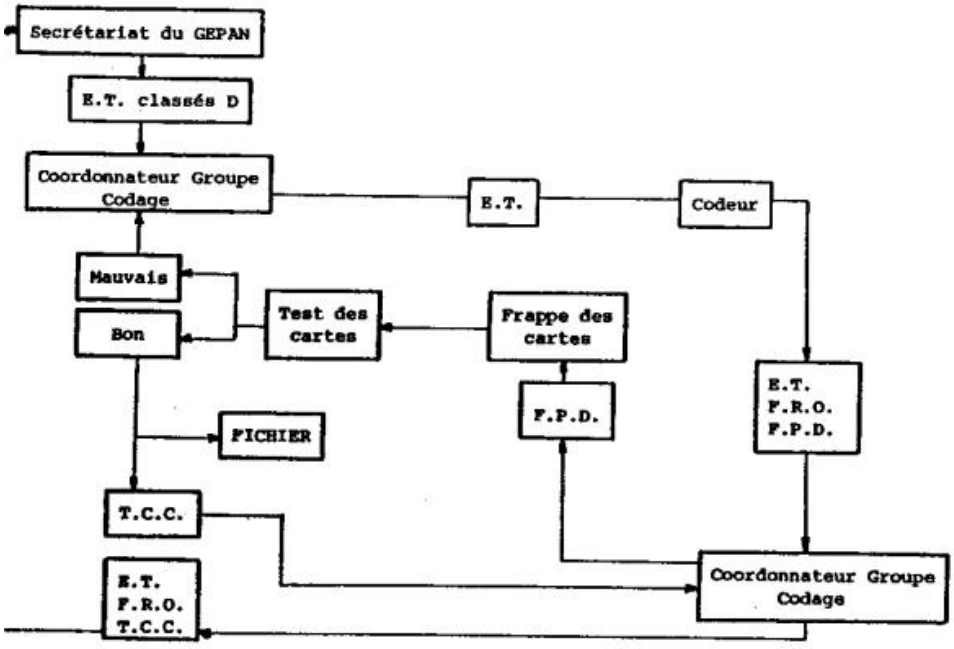
0. AGRICULTEURS EXPLOITANTS
   - 00. Agriculteurs exploitants
1. SALARIES AGRICOLES
   - 10. Salariés agricoles
2. PATRONS DE L'INDUSTRIE ET DU COMMERCE
   - 21. Industriels
   - 22. Artisans
   - 23. Patrons pêcheurs
   - 26. Gros commerçants
   - 27. Petits commerçants
3. PROFESSIONS LIBERALES ET CADRES SUPÉRIEURS
   - 34. Professions libérales
   - 32. Professeurs : professions littéraires et scientifiques
   - 33. Ingénieurs
   - 34. Cadres administratifs supérieurs
4. CADRES MOYENS
   - 41. Instituteurs s professions intellectuelles diverses
   - 42. Services médicaux et sociaux
   - 43. Techniciens
   - 44. Cadres administratifs moyens
5. EMPLOYES
   - 51. Employés de bureau
   - 52. Employés de commerce
6. OUVRIERS

- 60. Contremaîtres
- 61. Ouvriers qualifiés
- 63. Ouvriers spécialisés
- 65. Mineurs
- 66. Marine et pécheurs
- 67. Apprentis ouvriers
- 68. Manœuvres

7. PERSONNELS DE SERVICE
   - 70. Gens de saison
   - Femmes de ménage
   - Autres personnels de service
8. AUTRES CATÉGORIES
   - 80. Artistes
   - 81 : Clergé
   - 82. Armée et police
9. PERSONNES NON ACTIVES
   - 91. Étudiants et élèves de 17 ans et plus
   - 92. Militaires du contingent
   - 93. Anciens agriculteurs
   - 94. Retirés des affaires
   - 95. Retraités du secteur public
   - 96. Anciens salariés du secteur privé
   - 97. Autres personnes non actives de moins de 17 ans
   - 98. Autres personnes non actives de 17 à 64 ans
   - 99. Autres personnes non actives de 65 ans et plus

REMARQUES :
Les ensembles chiffrés par des numéros à deux chiffres sont *les catégories socioprofessionnelles*, les ensembles chiffrés par des numéros à un chiffre sont désignés sous le nos "groupes" dans le présent code.
Les noms des catégories et des groupes sont partiellement conventionnels, voir les paragraphes qui leur sont consacrés.

SCHÉMA DE CODAGE



LEGENDE :
E.T. : Enquête-témoignage
F.R.O. : Fiche Résumé d'Observation
F.P.D. : Feuille de Préparation de Données
T.C.C. : Test de Contrôle des Cartes
MARCHE A SUIVRE

Les membres du Groupe CODAGE trouveront auprès du coordonnateur :

- les enquêtes-témoignages (rapports de Gendarmerie, par exemple),
- les fiches résumé d'observation,
- les feuilles de préparation des données.

Après codage, c'est à dire remplissage des F.R.O. et des F.P.D., ils les ramèneront au coordonnateur qui se chargera d'assurer la frappe des cartes et de vérifier la conformité des cartes frappées.

En cas de non conformité, l'E.T. sera redistribuée aux codeurs pour un nouveau codage avec, éventuellement, une recherche et une discussion sur la non-conformité.

En cas de conformité, la carte sert à introduire le témoignage dans le fichier. De plus, le coordonnateur retourne au Secrétariat du GEPAN les E.T., les F.R.O. et les F.P.D. associées.

Les témoignages à coder sont ceux qui ont été classés deux fois en "D". Tout autre classement (D + C, par exemple) entraînera un retour au Secrétariat du GEPAN, par l'intermédiaire du coordonnateur.

CHAPITRE 3

RAPPORT D'ÉTUDES STATISTIQUES

Note technique faite sur proposition du Pr. DURANDIN et de Mme ASKEVIS
Travaux développés par Paul LEGENDRE.

RESUME

Un ensemble de 106 rapports d'observation classés dans la rubrique D formée *des rapports pour lesquels il n'est pas possible, malgré les détails décrits, d'identifier le phénomène observé*, a servi de base à la constitution d'une première version du fichier national d'observation. Ces observations ont eu lieu, dans leur majorité, durant les années 1974 à 1977. Seulement 9 rapports sont antérieurs à cette période. Les rapports proviennent tous de la Gendarmerie Nationale.

A partir des variables psycho-sociologiques codées dans le fichier, nous avons tenté de faire une analyse statistique. Cette note en présente les résultats.

1. LES VARIABLES PSYCHO-SOCIALES

Dans le code retenu (annexe 21 du rapport de juin 1978 au Conseil Scientifique), figurent les variables suivantes :

- département du lieu d'observation,
- nombre de témoins,
- profession du témoin principal
- âge du témoin principal.

Le dépouillement relatif à ces variables fournit les résultats consignée dans les paragraphes ci-dessous (le codage utilisé correspond à celui décrit clans la note n° 143/CT/GEPAN du 2 juin 1978).

1.1. DÉPARTEMENT DU LIEU D'OBSERVATION

Nbre de rapports	Département (code postal)	Effectif (Nbre Dépt)	Effectif théorique (*)
5	59.	1	10,5
4	21.54.76.82.	4	
3	13.28.34.38.45.51.85	7	
2	02.08.11.17.24.33.39.47.52.55.60.62. 68.71.80.83	16	21,5
1	04.07.09.10.15.16.19.23.25.26.29.30.32.35.42.46.50. 53.57.61.64.66.69.72.73.79.84.86.88.89.	30	39
0	01.03.05.06.12.14.18.20A.20B.22.27.31.36.37.40.41. 43.44.48.49.56.58.63.67.70.74.77.81.87.90.75.78.91. 92.93.94.95.65.	37	35

Plus deux rapporta provenant des territoires d'outre-mer français.
(*) : distribution aléatoire avec même probabilité pour chaque département.

La distribution de l'effectif par nombre de rapports ne diffère pas significativement de la distribution théorique correspondant à l'hypothèse d'une distribution uniforme (X^2 = 3,81 avec 3 d.d.l).

1.2. NOMBRE DE TEMOINS

L'histogramme est :

Nombre de témoins	Nombre de rapports
1	21
2	28
3	23
4 - 5	17
5	15
Non mentionné	2

L'effectif de la classe "1 seul témoin" est vraisemblablement sous-représenté dans la mesure où la présence de plusieurs témoins peut être un facteur incitant les "experts" à placer une telle observation dans la catégorie D. On retrouve cependant l'estimation de POHER (rapport du Conseil Scientifique de décembre 77) :

- "70 % des observations ont au moins 2 témoins et plus de 50 % des observations ont au moins 3 témoins ".

1.3..PROFESSION DU TEMOIN PRINCIPAL

Compte tenu du faible effectif global, nous n'avons retenu que la classification par groupes socio-professionnels (1er chiffre de la classification INSEE).

Groupe SOCIO-PROFESSIONNEL DU TÉMOIN PRINCIPAL	EFFECTIF
0 - Agriculteurs exploitants	12
1 - Salariés agricoles	2
2 - Patrons de l'industrie et du commerce	9
3 - Professions libérales et cadres supérieurs	4
4 - Cadres moyens	7
5 - Employés	6
6 - Ouvriers	24
7 - Personnel de service	7
8 - Autres catégories (en fait : Armées - Police)	11
9 - Personnes non actives	21
Non mentionné	3

Il est possible de détailler le groupe 6 :

6 - Ouvriers :	24
60. Contremaîtres	4
61. Ouvriers qualifiés	7
63. Ouvriers spécialisés	10
67. Apprentis ouvriers	3
68. Manœuvres	-

Il est possible de détailler le groupe 9 :

9 - Personnes non actives :	**21**
91. Etudiants et élèves de 17 ans et plus	1
92. Militaires du contingent	2
93. Anciens agriculteurs	-
94. Retirés des affaires	-
95. Retraités du secteur public	5
96. Anciens salariés du secteur privé	-
97. Autre personnes non actives de – 17 ans	6
98. Autre personnes non actives de 17 à 64 ans	4
97. Autre personnes non actives de + 64 ans	3

Ces données permettent de reprendre la conclusion de C. POHER :

- *"un très large éventail de professions existe parmi les témoins".*

On peut cependant préciser quelques points en comparant cet effectif à l'effectif théorique obtenu à partir de l'importance numérique de chaque groupe. Nous ne disposons de ces données que pour les catégories 0 à 8 (référence : recensement 1975).

Groupe socio-prof.	Nombre de personnes	Effectif théorique de rapports	Effectif constaté
0 - 1	2 026 315	7.6	14
2	1 708 925	6.4	9
3	1 459 2B5	5.5	1
4	2 764 950	10.4	7
5	3 841 255	14.5	6
6	8 207 165	30.9	24
7	1 242 935	4.7	7
8 (*)	357 980	2	11

(*) armée et police exclusivement.

Deux groupes sont sur-représentés :

- agriculteurs exploitants et salariés agricoles,
- armée - police,

et un groupe sous-représenté :

- employés.

Les différences sont significatives au seuil 0.025. La sur-représentation du groupe "armée-police" peut s'expliquer par le fait que nous travaillons exclusivement sur des rapports de gendarmerie et que les personnes de ce groupe sont plus prêtes à aller témoigner. A moins que cela ne résulte de l'expertise : cette catégorie apparaissant aux experts particulièrement crédible, encore que cet indice n'intervienne, en principe, qu'après la classification en A, B, C ou D.

La sur-représentation des agriculteurs et la sous-représentation des employés sont, semble-t-il, la conséquence d'un même phénomène : les observations d'OVNI sont faites à la campagne. Les agriculteurs, personnages de la campagne, sont beaucoup plus "favorisés" que les employés, personnages citadins.

1.4. AGE DU TEMOIN PRINCIPAL

CLASSE	AGE DU TEMOIN	EFFECTIF
E	enfant de 0 à 13 ans	3
J	adolescent de 14 à 20 ans	14
A	adulte de 21 à 59 ans	73
V	vieillard de 60 ans et plus	12
	information non disponible	4

Dans la très grande majorité des cas, le témoin principal est un adulte. Le faible effectif des classes E et J, résulte vraisemblablement de l'une ou l'autre des causes suivantes :

- l'expertise : les enfants sont considérés comme peu crédibles !
- en cas de témoignages multiples, le témoin principal est de préférence un adulte,
- "*aller voir les gendarmes*" pour raconter son histoire est plus difficile pour un enfant.

REMARQUE :

La nouvelle version des règles de codage permettra d'étudier plus précisément cette variable puisque l'âge figurera, explicitement et non à l'aide d'une classification a priori.

2. LES CORRÉLATIONS AVEC LES PHÉNOMENES OBSERVES :

Comme l'avait suggéré Mme ASKEVIS (lettre à C. POHER du 12.7.78), nous avons étudié les corrélations entre les variables psychosociales et aussi des paramètres décrivant l'observation :

- le type de phénomène (établi par les experts),
- la forme de "l'objet".

2.1. LE TYPE DE PHENOMENE OBSERVE

Sur l'ensemble des 106 cas analysés, on trouve :

- 46 cas : "lumière nocturne"
- 6 cas : "disque diurne"
- 4 cas : "observation instrumentale"
- 22 cas : "observation rapprochée type 1" (< 200 m)
- 25 cas : "observation rapprochée type 2" (traces)
- 3 cas : "observation rapprochée type 3" (avec débarquement).

Compte tenu du faible effectif, nous regroupons les trois premières catégories formant ainsi la catégorie C_1. La catégorie C_2 correspond exclusivement aux observations rapprochées de type 1 et la catégorie C_3 regroupe les deux dernières.

C1	C2	C3
56	22	28
53 %	21 %	26%

2.1.1. Relation avec la profession du témoin principal

GROUPE SOCIO-PROFESSIONNEL	C_1	C_2	C_3
0 -1. Agriculteurs exploitants et salariés agricoles	5 (35%)	2 (14s)	7 (50%)
2. Patrons de l'industrie et du commerce	4 (44%)	3 (33%)	2 (22%)
3. Professions libérales et cadres supérieurs	3 (75%)	1 (25%)	-
4. Cadres moyens	4 (57%)	3 (43%)	-
5. Employés	3 (50%)	-	3 (50%)
6. ouvriers	11 (46%)	8 (33%)	5 (21%)
7. Personnels de service	5 (71%)	1 (14%)	1 (14%)
8. Armée-police	6 (55%)	2 (18%)	3 (27%)
9. Personnes non actives	13 (62%)	2 (10%)	6 (29%)

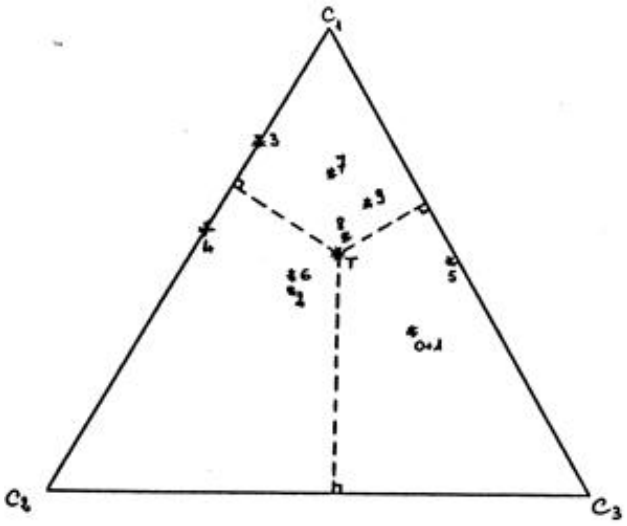
Compte tenu des petits effectifs pour chaque profession, les différences par rapport à une répartition homogène ne sont pas significatives. On peut cependant représenter géométriquement ces résultats pour faire apparaître les points suivants :

- les agriculteurs exploitants et salariés agricoles (0 + 1) ainsi que les employés se distinguent par une forte proportion de type C_3, les premiers l'emportant sur les seconds en ce qui concerne le type C_2 ;

- les professions libérales et les cadres moyens se distinguent an contraire par l'absence de type C_3.

REPRESENTATION GÉOMÉTRIQUE DE LA RÉPARTITION
DES TYPES D'OBSERVATION PAR PROFESSION

(La distance à l'axe $C_2\,C_3$ est proportionnelle à la fréquence de cas de type C_1 pour la catégorie représentée).



2.1.2. Relation avec l'âge du témoin principal

CLASSE	AGE OU TÉMOIN PRINCIPAL	C_1	C_2	C_3
E	de 0 à 13 ans	-	-	3 (100 %)
J	de 14 à 20 ans	10 (71%)	2 (14%)	2 (14%)
A	de 21 à 59 ans	36 (49%)	19 (26%)	18 (25%)
V	+ de 60 ans	7 (58%)	1 (8%)	4 (33%)

Si l'on excepte la classe E, dont l'effectif est très faible, on n'observe aucune variation significative de la répartition des types d'observation entre les différentes classes d'âge. L'âge du témoin principal et le type d'observation apparaissent indépendants.

2.1.3. Relation avec le nombre de témoins

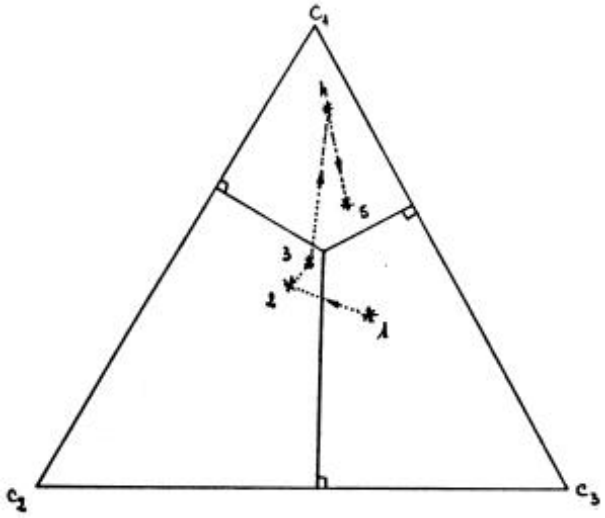
NOMBRE DE TÉMOINS	C_1	C_2	C_3
1	8 (38%)	4 (19%)	9 (43%)
2	12 (13%)	9 (32%)	7 (25%)
3	11 (48%)	6 (26%)	6 (26%)
4 - 5	14 (82%)	1 (6%)	2 (12%)
5	9 (60%)	2 (13%)	4 (27%)

Pour chaque modalité du nombre de témoins, les différences par rapport à la répartition globale ne sont pas significatives. Toutefois, apparaissent les phénomènes suivants :

- pour les quatre premières classes, la fréquence des rencontres (types C_2 et C_3) diminue. En effet :
  - 1 témoin 62 %
  - 2 témoins 57 %
  - 3 témoins 52%
  - 2-3 témoins 55%
  - 4-5 témoins18 %
  - + de 3 témoins 28%
  - + de 5 témoins 40 %

- parallèlement à cette évolution, on observe également une décroissance de la fréquence de la classe C_3 :
  - 1 témoin 42 %
  - 2-3 témoins 25,5 %
  - + de 3 témoins 19 %

REPRÉSENTATION GÉOMÉTRIQUE DE LA RÉPARTITION
DES TYPES D'OBSERVATION SUIVANT LE NOMBRE DE TÉMOINS



Autrement dit, "l'étrangeté" des observations diminue avec le nombre de témoins. Trois types d'explication peuvent être proposes :

- lors de l'observation d'un phénomène physique, le nombre de témoins est proportionnel à la distance d'où on peut voir le phénomène. Les rencontres rapprochées ont donc naturellement un nombre moindre de témoins ;

- la multiplicité des témoins est un facteur limitant l'écart entre la réalité observée et le récit de l'observation ;

- durant la sélection des cas, les experts ont retenu des cas avec un nombre restreint de témoins préférentiellement lorsque ceux-ci étaient "étranges".

2.2. LA FORME DE L'OBJET OBSERVE

L'étude de ce paramètre est plus délicat du fait de la grande variété de ses modalités, amplifiant ainsi les difficultés dues à la petitesse de l'échantillon. Par ailleurs, le codeur affecte une observation dans une classe de forme en fonction de son interprétation, ce qui impose une grande prudence pour l'utilisation de ces données.
Les résultats figurent ci-après, ne peuvent que suggérer des hypothèses. La répartition des forces est la suivante :

A - disque, soucoupe lenticulaire	10
B - ronde, circulaire, boule	32
C - cigare, cylindre, fusée	14
D - veuf, ovale, ovoïde, ballon de rugby	13
E - conique, triangulaire, trapézoïdal chapeau asiatique	8
H - soucoupe à coupole, chapeau canotier	7
K - ponctuelle, étoile, grosse planète	1
L - dôme, tasse, parachute, parapluie, seule de foin	2
N - méduse, champignon	1
P - croissant	2
Q - cigare accompagné de disques	1
* - information non disponible	7
. - non codable	8
TOTAL	**106**

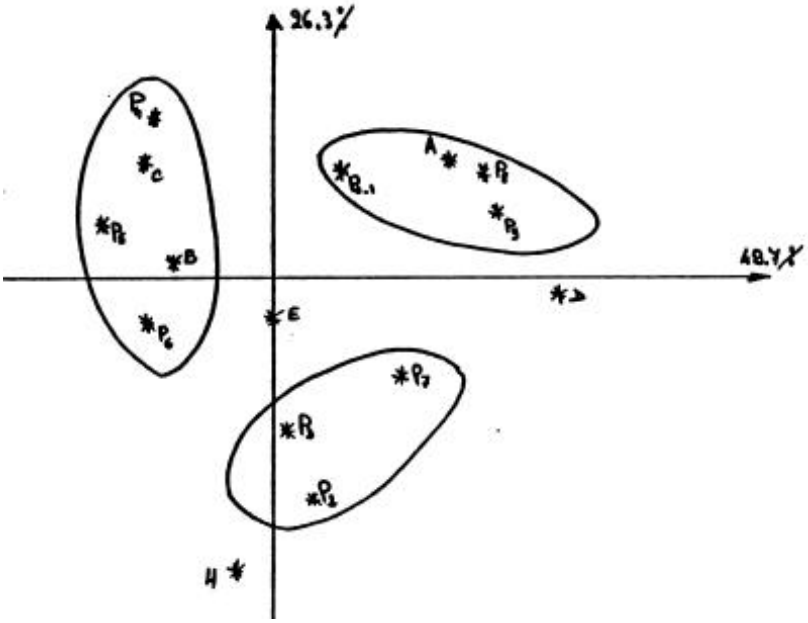
2.2.1. Relation avec la profession du témoin principal

Nous nous limiterons aux formes les plus représentées. Se reporter au tableau de la page suivante.

Afin de faire apparaître quelques faits de structures, nous avons procédé, malgré la faiblesse de l'effectif, à une analyse factorielle des correspondances sur ce tableau. (voir plus loin également).

CLASSE	GROUPE SOCIO-PROFESSIONNEL	A	B	C	D	E	H
0 - 1	Agriculteurs exploitants et salariés agricoles	2	3	2	2	1	-
2	Patrons de l'industrie et du commerce	-	3	-	2	1	2
3	Professions libérales et cadres supérieurs	1	1	-	-	-	1
4	Cadres moyens	1	3	2	-	-	-
5	Employés	-	4	1	-	1	-
6	Ouvriers	1	9	4	-	3	3
7	Personnel de service	-	2	1	3	-	1
8	Armée - police	-	4	2	-	-	-
9	Personnes non actives	4	3	1	6	2	-

Analyse factorielle des correspondances du tableau ci-dessus.
Représentation dans le plan des deux premiers facteurs :



On distingue un regroupement, en 3 catégories, des groupes socio-professionnels :

(1) "agriculteurs exploitants et salariés agricoles" avec "armée-police" et "personnes non actives". Ce groupe est fortement associé à la forme A et proche des formes D et E ;

(2) "cadres moyens" avec "employés" et "ouvriers". Ce groupe est associé aux gammes C et B et proche de E ;

(3) "patrons de l'industrie et du commerce" avec "professions libérales et cadres supérieurs" et "personnel de service". Ce groupe est proche de E, D et H (c'est le groupe le plus proche de H).

Remarquons que ce regroupement des groupes socio-professionnels n'est pas sans logique. Cependant, les liaisons avec le paramètre forme sont peu interprétables.

2.2.2. Relation avec l'âge du témoin principal

CLASSE	AGE DU TÉMOIN PRINCIPAL	A	B	C	D	E	H
E	13 ans	1	0	0	1	1	0
J	de 14 à 20 ans	0	6	1	3	1	1
A	de 21 à 59 ans	5	24	10	9	6	5
V	60 ans	2	2	3	0	0	1

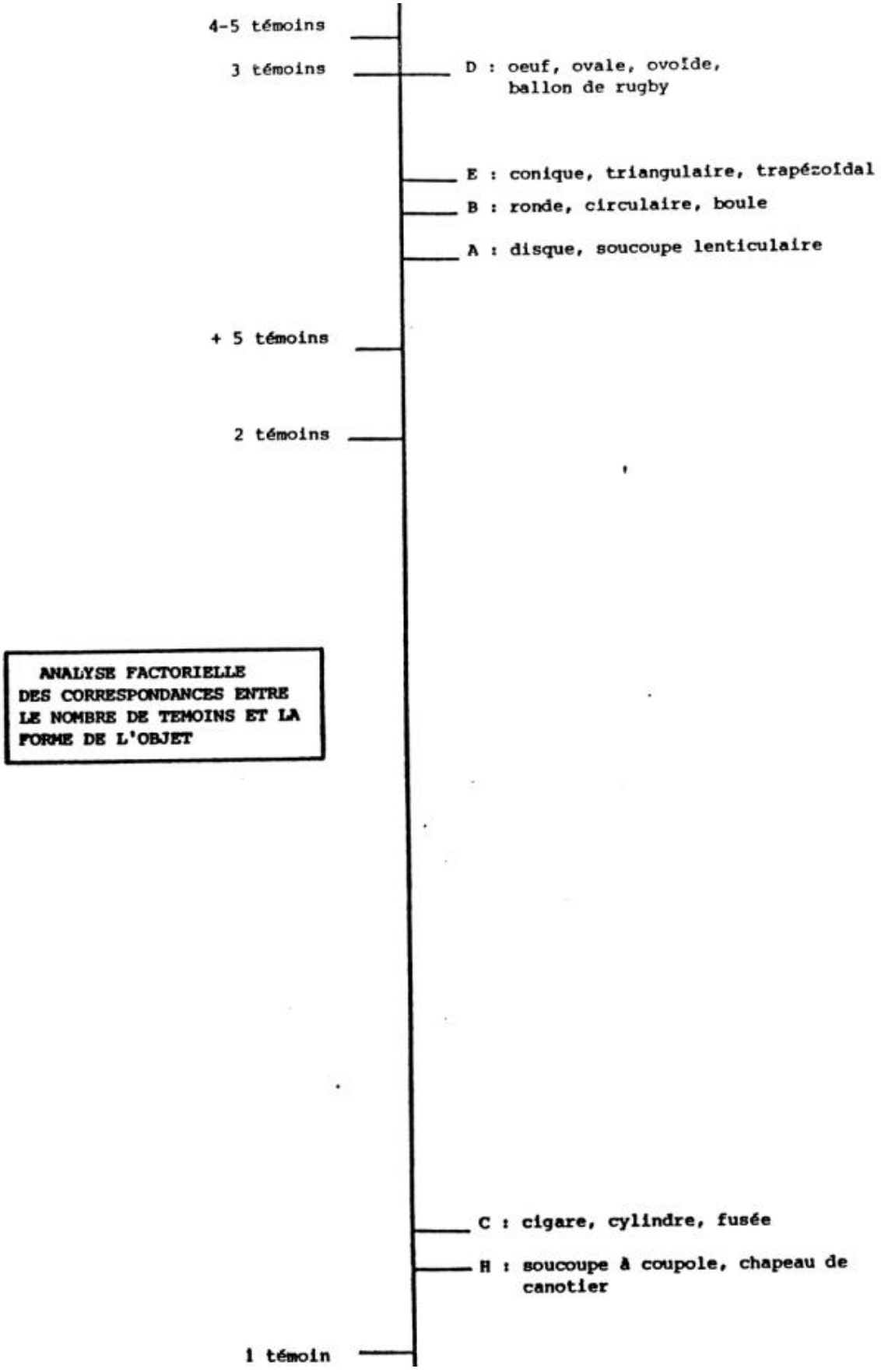
Pas de fait significatif.

2.2.3. Relation avec le nombre de témoins

FORMES	Nombre de témoins				
	1	**2**	**3**	**4 - 5**	**+ 5**
A	1	3	3	1	2
B	3	10	9	5	4
C	7	2	2	1	2
D	1	2	4	3	3
E	1	1	3	1	2
H	3	3	-	-	1
K	-	-	-	1	-
L	-	1	-	1	-
N	1	-	-	-	-
P	-	1	-	1	-
Q	-	-	-	1	-
*	1	2	1	1	1
.	3	3	1	1	0

Une analyse factorielle des correspondances, limitée aux seules formes A, B, C, D, E et H, conduit à extraire un premier facteur représentant 75 % de l'inertie totale du nuage. Une interprétation de ce facteur apparaît clairement : les quatre premières modalités se retrouvent ordonnées naturellement sur ces axes(*). Si l'on ne tient pas compte de la modalité "plus de 5 témoins", on peut dire que ce facteur représente le nombre de témoins.
(*) Ce phénomène est d'autant plus remarquable que l'analyse factorielle ne tient pas copte des relations d'ordre.

L'analyse factorielle est représentée ci après :



La part d'inertie expliquée (75 %) permet de conclure à une liaison entre le nombre de témoins et la forme de l'objet principal décrit.

Soumettons une hypothèse pour l'interprétation de cette liaison. Pour cela, nous distinguerons trois catégories de formes:

- formes que nous nommerons "interprétatives" : ce sont les formes suggérant un véhicule soit directement (fusée), soit indirectement (soucoupe à coupole = forme de véhicule extra terrestre popularisé par la science-fiction)

- formes que nous nommerons formes "anti-interprétatives" qui comme les précédentes, ne donnent une description que par l'intermédiaire de la référence à un objet sais sans que cet objet ne donne une interprétation de l'observation : ballon de rugby, neuf , meule de foin, etc…) ;

- enfin, formes géométriques : ronde, conique, trapézoïdale.

Bien que les classes de formes utilisées dans le codage ne soient pas homogènes vis à vie de ces catégories (ex. : la classe H comporte 1 élément des "formes interprétatives" : soucoupe à coupole, et 1 élément de "formes anti-interprétatives" : chapeau de canotier), nous proposons la classification suivante :

- formes interprétatives : C, H

- forces anti-interprétatives : D

- forces géométriques : A, B et E.

Notre hypothèse concernant la liaison : nombre de témoins/forme de l'objet décrit, peut s'exprimer alors :

- *"Lorsque le nombre de témoins est restreint, les témoins utilisent plus volontiers des formes interprétatives pour la description de leur observation".*

Il importe de remarquer qu'il s'agit là uniquement dune hypothèse suggérée par les données. Pour la confirmer, ou l'infirmer, il est nécessaire :

- de réexaminer la description fournie dans cheque rapport de gendarmerie afin de surmonter les difficultés dues au codage actuel,
- d' examiner le niveau d'indépendance des témoins .

3. CONCLUSION

L'analyse statistique dont nous venons de présenter les résultats a permis de mettre en lumière quelques faits.

Le plus remarquable est, à notre avis, la liaison entre le nombre de témoins et la forme de l'objet. Ce nombre de témoins est également lié au type de phénomène .

Il faut cependant rappeler que l'échantillon sur lequel porte cette analyse est très restreint. L'extension prochaine du fichier permettra de confirmer ou d'infirmer certaine résultats. Il nous semble toutefois que l'analyse effectuée permet de poser quelques hypothèses permettant de développer l'étude.

CHAPITRE 4

RÉFLEXIONS SUR LES PRINCIPES DU PRÉ-TRAITEMENT

Par Manuel JIMENEZ

EXPERTISE : Visite et opérations des experts

EXPERT : Personne apte à juger de quelque chose
(Larousse de poche)

Une expertise est donc un Jugement compétent. C'est à dire un classement d'un objet d'étude : l'étiquetage de cet objet avec une valeur d'une variable donnée (cette variable pouvant être simplement dichotomisée, ou être nominale (1)).

(1) Ou être composée de plusieurs variables qui présentent, ou pas, des valeurs caractéristiques. C'est le cas du "tableau clinique" ou du "profil psychologique".

Cela revient à dire que l'expertise est une mesure, et l'expert l'instrument de mesure. Une mesure doit répondre à des critères de fidélité et de validité :

- le critère de fidélité demande que deux mesures effectuées sur un même objet soient identiques (ou différer à l'intérieur d'une marge pré-établie), et indépendamment de toute variation extérieure à l'objet. Dans le cas de l'expert, cela signifie qu'un même objet doit être classé de la même façon par deux experts différents, ou par le même expert à deux moments différents ;

- le critère de validité exige que la mesure corresponds biens à une et une seule valeur de concept qu'elle est censée mesurer.

C'est à dire que le rapport entre la variable opérationnelle (définie par l'opération de mesure) et la variable théorique qu'on veut atteindre doit être le plus étroit possible.

Ce problème ne se pose pas lorsque la variable théorique et la variable opérationnelle ne font qu'une : lorsque le concept est défini par la mesure. C'est le cas de beaucoup de mesures physiques (par exemple, la température).

Par contre, lorsque le concept est "abstrait" (2), ce rapport ne peut être défini que par raisonnement logique à l'intérieur d'un cadre théorique donné. C'est le cas de la plupart des variables psychologiques (par exemple, l'intelligence) ; C'est aussi le cas, en principe de concepts tels que la crédibilité, l'intérêt, voire l'identification. De toutes façons, l'établissement de ce rapport passe par l'explicitation du concept dans une définition, qui est une définition opérationnelle, si non de la mesure, au moins du concept.

(2) "abstrait" défini, par négation, comme ce qui ne peut pas être atteint directement.

Ainsi, une façon commode de valider une mesure est de définir le concept, même s'il est abstrait, par la mesure elle-même (3). Ce qui est important c'est qu'une mesure constitue une information potentielle susceptible d'être utilisée pour un traitement ultérieur.

(3) C'est le cas de la boutade attribuée à BINET : "*L'intelligence, c'est ce que mesure mon test*" (d'intelligence).

Pour que ce traitement soit fait sans erreur, il est absolument nécessaire que :

a) le sujet qui mesure et le sujet qui traite l'information aient la même définition de concept théorique mesuré (ou censé être mesuré) ;

b) qu'ils soient d'accord sur les bases théoriques de la validité de la mesure, et sur son degré de fidélité.

Dire que cela équivaut à dire que l'information apportée par une mesure doit être accompagnée par l'information indiquant comment cette mesure a été faite. Cette deuxième information devenant ainsi partie pertinente de l'information apportée par la mesure.

DE L'IMPLICITE DANS LA MESURE

On pourrait penser que toutes ces explications ne sont pas nécessaires sans la limite où tous les sujets - ceux qui mesurent, ceux qui traitent la mesure, ceux qui lisent les traitements - s'adressent à une même définition implicite de la mesure et du concept mesuré.

Or, une innombrable quantité de travaux montrent l'extrême variabilité de la perception humaine qu'il s'agisse de la perception des données physiques ou de la perception sociale. En particulier, ces travaux montre que :

- même lorsque un sujet dispose d'un cadre implicite de mesure, ce cadre est soumis à des variations intra-individuelles non contrôlable, se traduisant par un aléatoire dans la mesure ;

- on constate, de même, des variations intra-individuelles considérables dans les mesures et les cadres correspondants, même lorsque ceux-ci sont explicités individuellement (4).

(4) Regarder, par exemple, les recherches de NOIZET et CAVERNI (1978) sur l'évaluation des lycéens, ou celles de ROSENTHAL et JACOBSON (1968) sur l'effet Pygmalion.

DE LA MESURE LORS DE L'EXPERTISE DES RAPPORTS D'OVNI

Très peu de choses sont explicites dans la mesure faite lors d'une expertise de rapports d'OVNI, qui conduit à classer ce rapport sans une des quatre catégories possibles : A, B, C ou D.

A notre connaissance, les seules explicitations existantes dans la littérature interne au GEPAN, font acte des définitions de ces quatre catégories :

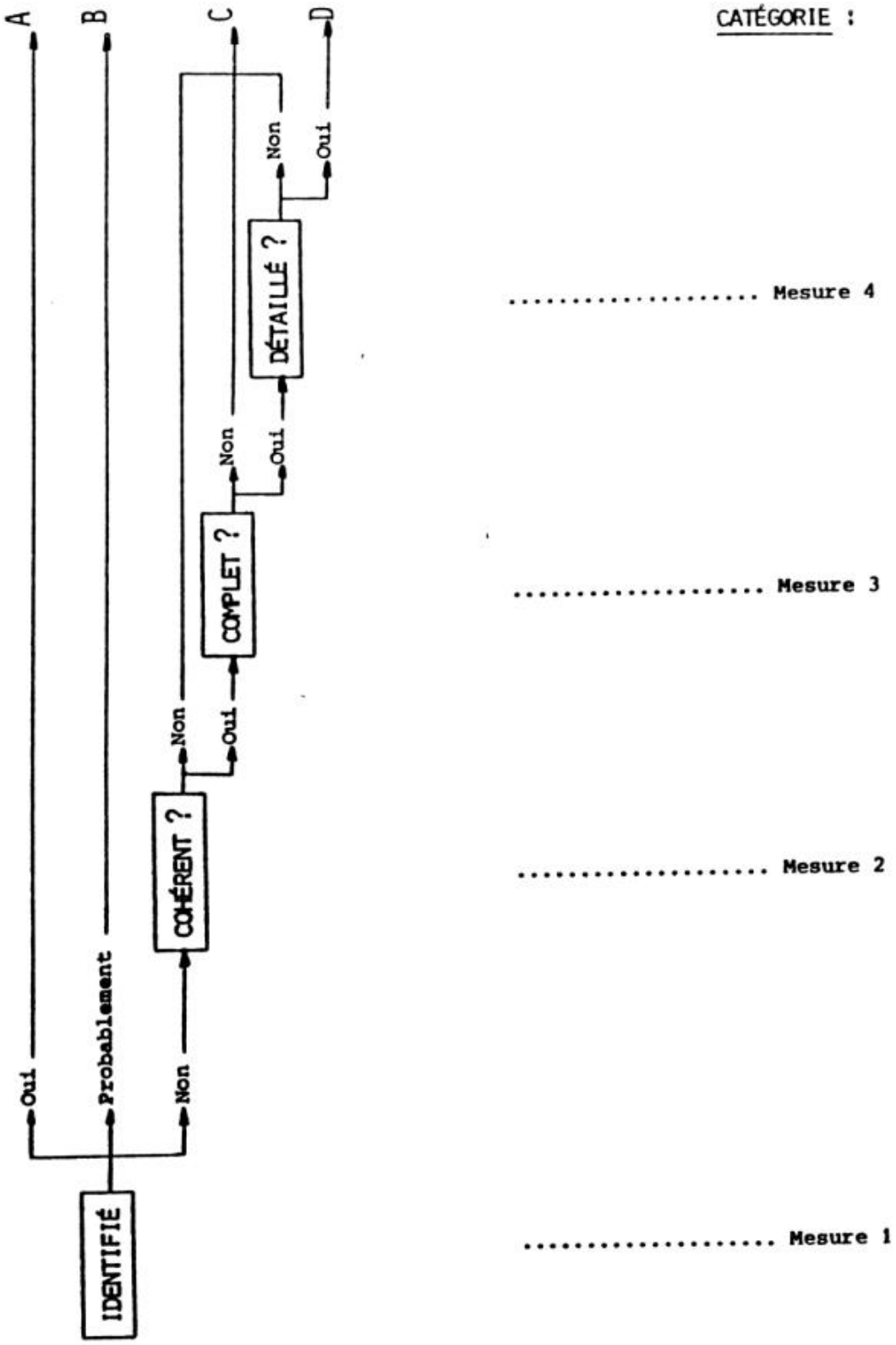
- A - phénomène identifié
- B - phénomène probablement identifié
- C - phénomène non-identifié mais le document manque d'intérêt (détails, cohésion,...)
- D - phénomène non-identifié et document cohérent, complet et détaillé.

Une analyse de ces définitions indique dans ces catégories un amalgame de plusieurs variables sous-jacentes :

- une variable identification du phénomène, ordinale, pouvant avoir trois valeurs : identifié, probablement identifié ou non-identifié ;

- trois variables ordinales dichotomisées : "détail", "cohésion" (ou "cohérence") et complétude desquelles on retient une combinaison ("cohérent, complet et détaillé") ; toutes les autres combinaisons (sept) sont considérées ensemble.

Cela semble pouvoir se traduire par la grille d'expertise suivante :



GRILLE D'EXPERTISE

Mais cela constitue uniquement une explicitation des valeurs possibles de quelques variables théoriques. En aucune manière, ne sont définies :

- ces variables,
- la façon comme elles et/ou leurs valeurs correspondantes sont mesurées.

Il résulte de cette analyse que, en absence de toute autre explicitation, les mesures faites sur ces variables, et le classement qui s'en suit, peuvent dépendre des variations infra et inter-personnelles inhérentes à toute grille d'analyse implicite humaine.

D'où le besoin...

QUELQUES PISTES

Deux mesures semblent avoir leur place lors de l'expertise d'un rapport d'OVNI :

- celle de l'information contenue dans le rapport,
- celle de l'identification de cette information.

Il semble difficile, lors d'une éventuelle mesure de l'information d'englober dans une seule échelle - dans une seule mesure - des estimations de la cohérence et de la quantité d'informations d'un rapport d'OVNI.

- Par cohérence, on entend l'absence de contradiction dans le rapport, au sens logico-linguistique.
- Par quantité d'informations on entend simplement, l'inverse de la probabilité d'apparition d'un message. Cela semble inclure la complétude et le "détail".

Un rapport peut être fort cohérent mais apporter peu d' informations et vice-versa. Mais, le manque de cohérence peut être considéré, en première analyse, comme une quantité d'informations supplémentaires . Dans la mesure où cette information supplémentaire n'est pas celle recherchée lors de la mesure de la quantité d'informations, l'expertise pourrait, pour l'éviter, procéder dans l'ordre :

- mesure de la cohérence,
- mesure de la quantité d'informations.

Une meilleure façon est de définir la "quantité d'informations" telle qu'elle exclut la cohérence (voir aussi à ce sujet le § 1 du chapitre 6).

L'OVNI étant par définition inconnu, on ne dispose pas de corpus de description permettant d'indiquer la probabilité d'apparition d'un rapport d'OVNI donné, ou d'une caractéristique d'OVNI donné. Mais, le GEPAN dispose d'une quantité non négligeable de descriptions, dont un certain nombre déjà codées, qui peuvent être utilisées comme matériau brut à l'élaboration de ce corpus.

La théorie de l'information doit permettre, à partir de ce corpus, la quantification univoque de l'information apportée par un nouveau rapport, en fonction de la probabilité d'apparition des caractéristiques y étant décrites et de l'agencement de ces caractéristiques.

Il est certainement plus difficile de quantifier la cohérence d'un rapport. Dans une première approximation, la cohérence peut se mesurer comme l'inverse du nombre de contradictions se trouvant dans un rapport. Une deuxième approximation devrait pondérer cette note en relation avec l'information brute du rapport : nombre de mots, ou de lignes, ou de caractéristiques (5) , etc.

(5) Il est logique de penser qua la probabilité, pour un témoin moyen, de se contredire croit avec la longueur de son rapport ....

- Cette information brute peut, le cas échéant, être considérée comme une toute première estimation d'une "quantité d'informations" du rapport (6).

(6) Il est nécessaire dans ce cas, de reprendre la définition de la "quantité d'informations" apportée par un rapport d'OVNI.

Mais, une bonne mesure - une bonne définition - de la cohérence d'un rapport doit tenir compte de la logique des contradictions, en fonction par exemple, de leurs probabilités d'apparition dans des descriptions d'ordre commun. Certaines contradictions sort certainement plus courantes - plus "normales" - que d'autres.

La mesure de l'identification de l'information continue dans un rapport d'OVNI semble fort délicate. Elle doit tenir compte en même temps, de la quantité d' informations externes - des OVI ou plutôt des PAI - dont l'expert dispose et du degré de relation entre les caractéristiques de l'OVNI expertisé et celles des OVI. Mais, elle doit aussi être pondérée par rapport à l'information contenue dans le rapport d'OVNI.

Autrement dit, une note brute de l'identification indique le nombre brut des caractéristiques de l'OVNI qui se retrouvent dans un OVI (y inclus les coordonnées Spatio-temporelles). Ensuite, cette note doit être pondérée en fonction de la quantité d'informations contenue dans le rapport, de la quantité d'informations extérieures et de l'importance relative des caractéristiques retrouvées.

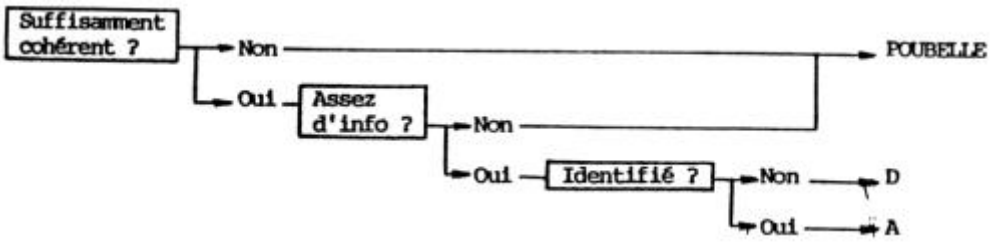
La fabrication de ces instruments de mesure n'est pas simple. Elle impose en même temps, un travail théorique (cf. la théorie de l'information) et un travail statistique à partir des rapports déjà codés, des informations des éventuels OVI.

Elle permet de créer, cependant, trois échelles :

- de cohérence,
- d'information,
- d'identification,

où on peut voter de façon univoque éventuellement chaque rapport d'OVNI.

A partir de là, l'expertise pourrait procéder comme suit :



En tout cas, un rapport classé de cette façon A ou D, dispose après expertise de trois notes objectives :

- cohérence,
- information,
- identification.

LIMITES

Cette démarche n'est que le fruit d'une réflexion rapide. Elle est destinée à subir des critiques, autant d'ordre théorique que méthodologique.

Elle ne prétend pas être la seule à suivre, pas même une parmi les démarches possibles. Elle est plutôt une illustration, certainement trop rapide, d'un cheminement qui se veut objectif avec le seul but d'ajouter le moins de bruit possible à l'information à traiter.

C'est le premier souci à avoir, au cours de l'expertise.

CHAPITRE 5

ETUDE DES PROBLÈMES LIÉS A LA CRÉATION D'UN FICHIER INFORMATIQUE

Par JP. ROSPARS

"*Ce n'est qu'au jour où (...) les observations seront extraites des archives poussiéreuses ; où la négligence et l'ignorance. Les ont enfouies que la Science sera en mesure d'analyser le phénomène OVNI*".

Jacques VALLEE (1)

RÉSUMÉ

L'existence des témoins, des traces alléguées et des documents qui en émanent ou qui s'y rapportent, constitue un phénomène OVNI empirique incontestable. Pour en déterminer les causes, il convient de rassembler des données sur les observables disponibles par des méthodes rigoureuses. Ce recueil est matériellement difficile pour les témoins et les traces, il l'est moins pour les documents dont le nombre élevé exige le recours à l'informatique.

La construction d'un fichier informatique pose essentiellement des problèmes ufologiques. Le premier est celui de l'unité de base du fichier (entrée) qui peut être le phénomène allégué, le témoignage ou le document. Ce dernier choix n'impose sauf exception. Le second est celui de la liste des caractéristiques à coder (variable) : elle doit être déterminée en tenant compte des phénomènes eux-mêmes mais tels qu'ils apparaissent en moyenne dans les documents, et des fonctions attendues du fichier. Ces fonctions qui ont chacune leurs exigences propres sont : la mise en ordre des documents, la recherche bibliographique et les études statistiques. Le troisième problème est celui de la création et de l'application pratique des codes (modalités des variables) Leur création est conditionnée par le caractère multidimensionnel des données et se fonde sur le concept d'échelle de mesures. Leur application introduit l'équation personnelle du codeur, qui peut être limitée, entre autre par l'emploi d'un livret de codage. Enfin, Il convient de tenir compte des fichiers existants, dont certaine sont extrêmement développés, et de l'expérience ainsi acquise.

En conclusion, une philosophie directive est proposée insistant sur la généralité du fichier tant du point de vue des documents traités que des applications possibles, et sur sa structure hiérarchique, tous les documents et toutes les variables n'ayant pas même importance. Cette philosophie appelle en contrepartie, le développement de fichiers plus spécialisés.

1. INTRODUCTION : L'ETUDE SCIENTIFIQUE DES DOCUMENTS OVNI

1.1. TENTATIVE DE DEFINITION DU PHENOMENE OVNI

On désigne communément sous le nom de "phénomène(s) OVNI" un ensemble disparate qui peut inclure un ou plusieurs des éléments suivants : des "objets volants non-identifiés", des témoins, des traces, des photos, des interprétations diverses, des groupes d'enquêtes privés, des répercussions dans les média, etc... Il est clair que ces divers éléments n'ont pas le même statut scientifique. Ainsi, l'OVNI lui-même n'est pas actuellement susceptible d'étude scientifique,

compte tenu de son existence problématique et de sa manifestation irreproductible. Il convient donc, avant toute chose, d'introduire une terminologie rigoureuse, adaptée à un usage scientifique, et apte à faire l'unanimité des personnes compétentes. Aussi proposons-nous la définition suivante en deux volets :

- le phénomène OVNI est constitué par l'existence dans de nombreux pays du monde et de manière continue depuis la fin de la seconde guerre mondiale, de personnes qui ont rapporté avoir vu des lumières ou des objets inhabituels dans le ciel, ou au sol ou près du sol. Ces témoignages décrivent parfois des objets solides d'aspect artificiel et sont accompagnés, de manière rare mais significative, d'enregistrements physiques (photo, écho radar, ...) ou de traces au sol attribuées au phénomène allégué ;

- nombre de ces témoignages, soumis à une analyse méthodique auprès des témoins et sur le site de l'observation alléguée, ne peuvent être attribués en toute rigueur ni à des affabulations ni à des mésinterprétations (*) de par la personnalité des témoins et la spécificité des données qu' ils fournissent. Ils demeurent donc inexpliqués non seulement pour les témoins mais également par les "experts".

(*) Nous proposons ce néologisme construit sur le modèle de mésalliance, mésentente, mévente, de préférence à "mauvaise interprétation" ou autres expressions similaires.

1.2. COMMENTAIRES DE LA DÉFINITION

L'existence de témoins (premier volet de la définition) est un fait tangible parfaitement établi. Son interprétation est l'objet d'une controverse passionnée depuis 30 ans. Les informations actuellement disponibles révèlent que cette existence ne peut être aisément expliquée (second volet) puisque certains témoins fournissent des témoignages articulés dont les meilleurs, comme les enquêtes du GEPAN l'ont montré, sont capables de résister à des tests sévères qui établissent au moins l'existence historique d'un évènement "anormal" et établiraient en toute autre circonstance (c'est-à-dire si le stimulus à l'origine du témoignage était connu) la réalité physique de l'observation. Il n'est dès lors pas exclu que le phénomène OVNI puisse constituer un phénomène nouveau : spécifique et irréductible.

L'expression "phénomène OVNI" présente de ce fait une certaine ambiguité puisqu'elle peut désigner :

- ou bien un fait objectif : l'existence de témoins, l'émission de rapports. C'est ce sens que nous avons retenu pour construire la définition qui précède. Pour éviter une confusion extrêmement préjudiciable, l'expérience le montre, à une réflexion claire sur ces questions, nous qualifierons ce fait de "phénomène OVNI empirique". L'emploi du singulier est justifié par le grand nombre de propriétés communes présentées par chacune des manifestations du phénomène (apparition d'un témoin) ;

- ou bien une explication possible du fait précédent, qu'on peut alors qualifier de "phénomène OVNI explicatif". L'emploi du singulier prête alors à discussion puisqu'il n'y a aucune raison de penser a priori qu'un seul phénomène est responsable du fait empirique.

Cette dernière remarque soulève une difficulté quant à la définition précise du phénomène OVNI empirique : doit-on ou non y inclure les témoignages expliqués par affabulation consciente ou inconsciente et par mésinterprétation ? Il est certes préférable, pour que l'attention soit centrée sur l'essentiel, de n'inclure que les témoignages peu ou pas réductibles. Toutefois, il est clair qu'une évaluation rigoureuse de chaque cas est et demeurera difficile pour des raisons matérielles. A

chaque instant, le phénomène empirique est donc un ensemble hétérogène comportant un noyau dont on sait positivement qu'il est très difficilement explicable et une périphérie dont on a déterminé l'explication précise ; entre les deux, un sous-ensemble se rattachant au noyau ou à la périphérie selon la quantité d'informations disponibles et la possibilité d'une explication. Pour des raisons méthodologiques, un examen comparatif de l'ensemble des cas expliqués ou non semble indispensable en l'état actuel de nos connaissances. Ainsi, considérons nous dans ce qui suit le phénomène empirique au sens large.

1.3. L'ETUDE SCIENTIFIQUE DU PHENOMENE

Les témoins, les témoignages, les "traces" de toute nature, constituent la matière première, les observables, sur lesquels le GEPAN doit travailler pour tenter d'établir les propriétés et, de là, la nature du ou des phénomènes qui engendrent les témoignages et ces traces allégués. En pratique, le GEPAN n'étudie pas un phénomène conjectural (tels que les "soucoupes volantes" par ex.) mais un phénomène réel qui tient à l'existence expérimentalement certaine de témoins et de "traces". Il convient donc de bien distinguer les phénomènes observables (témoins, traces, etc...) qui sont du domaine de l'observation et de l'expérimentation, observation et expérimentation destinées à recueillir des données aussi diverses, complètes que possible sur ces observables, et les phénomène explicatifs qui sont du domaine de l'hypothèse de travail, du modèle, de la théorie, ces diverses constructions rationnelles devant être confrontées aux donnée recueillies sur les observables pour infirmation ou confirmation.

Le recueil des données sur le phénomène OVNI empirique exige la mise au point d'une méthodologie adaptée à l'observable considérée qu'elle soit une trace, un témoin ou un document. En ceci l'étude du phénomène OVNI n'est en rien différente de celle pratiquée dans toute autre discipline scientifique. La science ne se définit pas par son objet mais par ses méthodes : il n'y a pas d'objets d'études indignes de la science, il n'y a que des méthodes indignes d'elle. Les études au GEPAN n'échappent pas à l'activité scientifique par le fait qu'elles s'appuient principalement sur des témoins et sur des témoignages oraux et écrits, car ce sont là des observables disponibles et qu'on n'a le choix qu'entre les utiliser ou ne rien faire. Elles y échapperaient par contre si les données étaient recueillies de manière hâtive et approximative, si aucun effort n'était fait en vue de l'évaluation critique ou de l'amélioration des méthodes utilisées ou si les analystes se laissaient entraîner par leurs idées a priori.

Le premier travail du GEPAN, celui-là même qui peut suffire à justifier son existence d'un point de vue scientifique, est précisément de développer des méthodes aptes à permettre le recueil de données ufologiques et de recueillir effectivement ces données à partir des observables disponibles (1ère phase de la méthode expérimentale) Ensuite, mais ensuite seulement, dans l'ordre logique mais aussi dans l'ordre des priorités, il lui reviendra à proposer des hypothèses de travail de portées plus ou joins grandes destinées à rendre compte de tel ou tel aspect du phénomène (2ème phase) et à les tester au moyen des données recueillies (3ème phase). Si cette démarche est suivie, on pourra dire que le GEPAN fait oeuvre scientifique et que le phénomène OVNI est scientifiquement étudié (2).

Nous nous proposons ici d'étudier les problèmes méthodologiques posés par la 1ère phase, celle du recueil des données, et par une seule des catégories d'observables, celle des documents.

1.4. LE RECUEIL DES DONNEES

Le phénomène OVNI empirique peut être l'objet d'une étude direct qui consiste pour les chercheurs du GEPAN à recueillir des données auprès des témoins eux-mêmes sur le site de l'observation. Ces chercheurs peuvent alors analyser les "traces" observables qui se rapprochent

le plus de ce sur quoi travaillent les chercheurs des sciences physiques et biologiques. Ils peuvent également, étudier des témoins et ce de plusieurs manières différentes : non seulement en leur demandant de conter un récit et de répondre à des questions (témoignage au sens strict) mais aussi en faisant usage de techniques non verbales (choix d'un couleur dans un teintier, détermination d'une direction à l'aide d'un théodolite etc...) dont les limitations sont très certainement différentes des techniques fondées sur l'expression orale.

Le GEPAN doit évidemment travailler autant que possible sur des traces et des témoins. Cependant, cette approche est limitée par deux facteurs le premier est le temps écoulé depuis l'observation qui affects les souvenirs des témoins encore vivants et qui rend parfois difficile la recherche de leurs adresses actuelles ; le second est le nombre même de ces témoins qui, compte tenu du coût des enquêtes, impose d'en réserver l'usage à un faible échantillon. Force est donc de recourir à une étude indirecte qui utilise les enquêtes faites par les journalistes, les enquêteurs privés et les gendarmes. Les documents ainsi mis en circulation sont des sources de données dont la qualité est en principe moindre que celles que le GEPAN aurait pu obtenir directement mais qui ont le mérite essentiel d'exister, d'avoir été élaboré indépendamment du GEPAN et de ne lui coûter que le prix de leur collecte.

L'étude des documents ne s'impose d'ailleurs pas seulement pour des raisons pratiques qui viennent d'être dites, qui en feraient une sorte de moindre mal, mais de mal tout de même. En fait, il convient de les étudier également en tant que phénomène propre. Si le document peut être analysé comme le sous-produit de l'exposition du témoin à un phénomène exceptionnel, il peut l'être aussi et indépendamment comme expression d'un phénomène de société qui possède sa problématique propre. L'existence des documents appelle ces développements sociologiques et historiques.

1.5. L'ETUDE DES DOCUMENTS

Les documents ufologiques émis en France depuis la fin de la guerre, sont extrêmement nombreux : probablement plusieurs dizaines de milliers. Ils sont dispersés : articles de journaux (essentiellement de la presse régionale), revues d'amateurs (souvent de faible tirage, d'existence éphémère et de collecte difficile), livres de toutes natures, rapports officiels. Ils sont redondants : dans biens des cas, une même observation donne lieu à plusieurs documents qui, si l'observation a été l'objet d'une certaine publicité, forment un réseau touffu.

Ces documents sont d'origine très hétérogène : information de seconde main, interview plus ou moins rapide de journalistes, enquêtes d'amateurs plus ou moins objectifs et compétents, sporadiques jusqu'en 1965 environ, plus nombreuses depuis et enquêtes officielles négligentes ou approfondies, sporadiques jusqu'en 1975, et nombreuses depuis. Les informations y sont consignées de manière non systématique : il est peu de questions auxquelles tous les documents donnent réponse.

Enfin, les évènements décrits sont eux-mêmes très hétérogènes par leur nature : du point lumineux dans le ciel à l'objet posé dans un champ, et par leur complexité qui dépend du nombre de témoins interrogés, du nombre de phases observées et de l'intensité de l'interaction phénomène/témoins.

L'étude des documents implique donc la résolution de trois types de problèmes :

- la collecte de documents,
- leur classement,
- le recueil des données à partir des documents collectés.

Nous ne considérerons ici que les deux derniers problèmes et surtout le troisième poursuivant et explicitant ainsi une étude antérieure (3) .

Le nombre et la dispersion des documents exigent sans conteste le recours à l'informatique. Cet outil peut être utilisé de deux manières complémentaires : soit pour effectuer le recensement et le classement logique des documents existants, soit pour étudier le contenu informatif de ces documents. Il s'agit dans les deux cas, de convertir des données relatives au document ou contenues dans le document, en informations stockées dans un fichier informatique. Cette opération conserve le nombre, la redondance et l'hétérogénéité propre aux documents (ou aux phénomènes) mais réduit leur caractère non-systématique et supprime leur dispersion.

L'objectif à atteindre est donc clairement fixé. Il reste à préciser quels documents utiliser et quelles données coder, à expliciter les méthodes de recueil des données, c'est à dire, compte tenu de l'option informatique choisie, les méthodes de codage, à prévoir leur amélioration incessante à l'avenir et à former en conséquence des codeurs expérimentés. Tous ces points de procédure ont déjà été mis en oeuvre au GEPAN avec une attention toute particulier pour la collecte, le collage et le traitement des rapports de gendarmerie. Ces rapports ne constituant qu'une petite fraction (mais de bonne qualité) de l'ensemble des documents disponibles, il convient d'étudier l'élargissement de ces procédures à l'ensemble des documents disponibles et de mieux préciser à cette occasion, les objectifs visés à travers la création d'un fichier informatique. Tels sont les points que nous nous proposons de discuter dans ce qui suit.

2. NATURE DES ENTREES (STRUCTURES VERTICALE)

Le premier problème posé par la constitution d'un fichier informatique est relatif à sa structure verticale. Un fichier est constitué d'unités de base ou entrées. Quelle doit être la nature de ces entrées ?

Pour répondre à cette question, il est indispensable d'expliciter les étapes de la génération d'un document ufologique, et de voir quels sont les problèmes posés aux analystes par chacune de ces étapes.

2.1. PROCESSUS DE GENERATION D' UN DOCUMENT

Supposons pour fixer les idées, qu'un phénomène physique quelconque soit à l'origine des observations et des traces alléguées: La première phase du processus est alors la manifestation d'un tel phénomène au-dessus dune zone S à partir d'un instant t.

Ce phénomène est observé par des témoins isolés ou en groupe. Si les témoins ont échangé des informations au moment de l'observation, ils sont dépendants. Si les témoins ont découvert et observé le phénomène sans échanger d'informations, ils sont indépendants. Bien entendu, il existe des transitions entre l'indépendance complète et la dépendance totale. Un échange d'information peut avoir lieu après la découverte indépendante voire après l'observation indépendante.

Les témoins peuvent communiquer leur observation sous forme d'un document écrit par eux-mêmes, ou la confier à un rapporteur quelconque qui en assure la rédaction. Un document primaire est ainsi engendré. On notera que les enregistrements d'interview sur bandes magnétiques, les films, les photos, les relevés peuvent également être considérés comme des documents primaires.

Tout document qui ne résulte pas d'un entretien direct avec un témoin ou qui fait preuve dune utilisation quelconque d'un document primaire (par citations, traduction, résumé, analyse, etc...) est un document secondaire.

D'après ce qui précède, une entrée du fichier peut donc être soit une manifestation du phénomène, soit une observation par un témoin ou un groupe de témoins dépendants, soit enfin un document faisant état d'un ou plusieurs témoignages.

2.2. DETERMINATION DE L'ENTREE

L'analyste peut-il prendre la manifestation du phénomène comme entrée du fichier ? C'est impossible pour deux raisons :

- la première, fondamentale, est qu' il est généralement difficile d'établir que différents témoins ou groupes de témoins indépendants et séparés dans l'espace ou le temps ont observé un seul et même phénomène. Le problème ne doit pas être supposé résolu avant la création du fichier. C'est au contraire le fichier préalablement constitué qui devrait permettre de le résoudre ;

- la seconde raison est d'ordre pratique : dans nombre d'études, les témoignages s'avèreront plus intéressants en eux-mêmes que le phénomène. Supposons par exemple, que l'analyse révèle qu'une série observations est due à une rentrée de satellite. Ce n'est évidemment pas la rentrée de satellite qui sera l'objet de nos études, mais le nombre et la répartition géographique des témoins, ainsi que les descriptions indépendantes qu'ils donneront du même évènement.

Les témoignages d'un témoin ou d'un groupe de témoins dépendants peut-il constituer une entrée ? Non encore, et pour deux raisons :

la première est que le(s) témoin(s) a(ont) pu donner lieu à plusieurs documents primaires qui peuvent être partiellement contradictoires. Or, il ne faut pas demander au codeur de confronter des documents et de choisir la version qui lui paraît la meilleure, car un tel choix pourrait se révéler subjectif. La révélation de ces contradictions sera au contraire l'une des tâches de l'analyse que l'on pourra effectuer à partir du fichier

la seconde raison est que le document primaire n'est pas toujours entre les mains du GEPAN : il faut donc se contenter de documents secondaires (du moins provisoirement).

Chaque entrée du fichier doit donc être constituée par un document primaire ou secondaire relatif à un témoin isolé ou à un groupe de témoins dépendants. Sont considérés comme documents, tout rapport, article, livre ou passage de livre signalant l'existence d'un groupe de témoins dépendants, quelles que soient par ailleurs la qualité ou la quantité des données fournies.

Toutefois, le principe "entrée = document relatif à un groupe de témoins dépendants", peut subir deux exceptions de sens contraire :

- la première concerne un document qui fait état de plusieurs observations par des témoins indépendants sans fournir de détails spécifiques, même sommaires, sur chaque observation. Un tel document peut être qualifié de "synthétique". Il constituera une entrée unique, mais cette particularité sera codée ;

- la seconde exception concerne un document très complet où chacun des témoins d'un groupe "dépendant" a été entendu séparément. Les observations peuvent alors différer suffisamment entre elles pour justifier une entrée séparée par témoin. Un codage synthétique peut être ultérieurement réalisé après une étude comparée.

On notera qu'en raison de ces deux exceptions, les entrées ne constituent pas un ensemble homogène. Aussi conviendra-t-il de prendre des précautions lors de l'exploitation statistique du fichier.

2.3. CONCLUSION : OBJECTIVITE ET GENERALITE DU FICHIER

Le fichier doit entrer tout document qui relate l'observation par un témoin ou un groups de témoins dépendants d'un phénomène considéré à tort ou à raison comme un OVNI soit par les témoins eux-mêmes, soit par les enquêteurs, soit par les commentateurs.

Cette clause est valable même s'il est probable ou certain que le rapport résulte d'une mésinterprétation ou d'une affabulation. Bien entendu, cette information devra être codée

Le codeur n'a pas pour tâche de confronter des documents distincts ni de décider si un rapport est cohérent ou non, ni d'établir si différents rapports décrivent ou non le même phénomène. Redisons le : tous ces problèmes doivent être laissés à l'analyse ultérieure du fichier.

Une telle procédure garantit l'objectivité et la généralité du fichier. Ainsi, les études comparatives et statistiques ne seront pas biaisées par une distinction nécessairement subjective au départ entre les "bons" et les "mauvais" cas d'OVNI. Toutefois, ces considérations n'impliquent :

- ni un quelconque désintérêt vie à via de l'expertise en profondeur, que peut mener le GEPAN sur tel ou tel cas spécifique,

- ni le maintien en toutes circonstances d'une valeur égale pour tous les documents. Une fois le fichier général constitué, rien n'empêchera un analyste de faire une discrimination des documents en fonction de la richesse ou de la crédibilité de leur information par exemple.

3. LE PROBLÈME DES VARIABLES (STRUCTURE HORIZONTALE) ("QUOI CODER" ?)

Une entrée est constituée d'unités descriptives ou variables qui concernent chacune un aspect du témoignage. Quels doivent être le nombre et la nature de ces variables ? Cette double question, relative à la structure horizontale du fichier, est beaucoup plus complexe que celle de sa structure verticale car elle dépend à la fois des documents à décrire (dont nous supposerons connus les traits majeurs) des documents disponibles (§ 1.1.) et des recherches à entreprendre à partir du fichier (§3.2.).

3.1. QUALITE DES DOCUMENTS ET NOMBRE DE VARIABLES

Combien de variables faut-il introduire pour chaque entrée ?
Le principal facteur à prendre en compte pour en décider est la qualité moyenne des documents disponibles. En effet,

- si la structure horizontale comportait peu de variables et de modalités par variable, le nombre d'entrées contenant des colonnes vierges (non codées) serait faible mais par

contre, le nombre de documents qui auraient pu donner lieu à un codage plus élaboré serait élevé. Une telle situation entraînerait donc en moyenne une perte d' information ;

- par contre, si la structure horizontale comportait beaucoup de variables et de modalités, un grand nombre d'entrées auraient beaucoup de colonnes non codées en raison de l'absence d'informations dans les documents. Le nombre de documents pour lesquels il y aurait beaucoup de variables non codées serait alors élevé. Une telle structure serait donc inutilement lourde dans la plupart des cas.

La meilleure solution consiste, nous semble-t-il, à prévoir une structure horizontale un peu plus large que celle qui correspond à la moyenne des documents. De cette manières, on pourra mesurer la qualité du document par le remplissage plus ou coins complet et précis de la structure et isoler de la masse les documents de qualité "supérieure" en nombre suffisamment restreint pour permettre une étude séparée. En un mot, il importe absolument d'éviter le fichier général fourre-tout nivelé par le bas.

Cette conception débouche sur l'idée dune structure hiérarchique du fichier :

- le fichier général est prévu pour coder tous les documents quels qu'ils soient. Les éventualités peu fréquentes ne pourront être codées en détail, malgré leur grand intérêt potentiel pour l'étude du phénomène. Il importe que le fichier général signale donc l'existence de ces éventualités. Une bonne structure horizontale doit impérativement mentionner ses propres insuffisances ;

- la conception ci-dessus rend nécessaire le développement de fichiers spécialisés dérivés du fichier général, de manière à permettre l'étude plus approfondie des documents qui le méritent. Il nous semble pat exemple, que le collage spécifique des cas de rencontres rapprochées s'impose en premier lieu.

Le développement de ces fichiers annexes ne pose pas de problème particulier. La plus sûre garantie d'efficacité clans ce domaine serait sans doute de laisser leur réalisation aux soins non d'un groupe de codeurs spécialisés, mais aux chercheurs intéressés qui souhaiteraient davantage d'informations sur certains documents. Un collage est d'autant mieux fait qu'il est réalisé par une seule personne et dans un but précis.

3.2. FONCTIONS DU FICHIER ET LISTS DES VARIABLES

Le fichier doit être construit en vue de remplir certaines fonctions. Il n'est pas un but en soi, il n'est qu'un moyen et, à oublier ce fait, on s'expose à de graves déconvenues. Le fichier doit donc permettre à lui seul, ou tout au moins faciliter, des études scientifiques sur le phénomène OVNI. On peut distinguer trois types de fonctions que doit remplir le fichier :

- mise en ordre des documents.
- recherches bibliographiques,
- recherches statistiques.

3.2.1. Fonction d'ordonnancement

- Bloc :

  - La première tâche à effectuer est de déterminer si deux documents diffèrents proviennent ou non du même groupe de témoins dépendants. L'ensemble des entrées

décrivant le même évènement mais fondées sur des sources différentes est appelé bloc (nomenclature de SANDERS). Tout bloc possède une structure soit chronologique (pour les documents primaires), soit hiérarchique (pour les documents secondaires qui ne peuvent être plus complets ou plus précis que les documents dont ils dérivent). Le problème est donc double : il faut d'abord assigner une entrée à un bloc puis hiérarchiser les entrées au sein d'un bloc. Ces deux opérations supposent la présence des variables suivantes :

- date et lieu de l'évènement,
- nom du ou des témoin(s) pour l'assignation,
- date de publication et source du document pour la hiérarchisation

- Super-bloc :
  - Il faut également pouvoir déterminer si deux blocs provenant de témoins indépendants correspondent ou non à l'observation du même phénomène. Ceci ne peut se faire que sur des critères plus ou moins conventionnels de proximité dans le temps et dans l'espace, et de similitude des descriptions. Cette opération est très Importante pour fournir une preuve directe de l'objectivité du phénomène, mais elle intervient également dans les études statistiques spatio-temporelles.

3.2.2. Fonction bibliographique

Le problème à résoudre est le suivant : l'analyste désire étudier les documents contenant certaines informations spécifiées. Il dispose donc d'une série de critères et se sert du fichier pour établir la liste des entrées qui répondent à ces critères. Le traitement bibliographique vise ainsi à mettre à la disposition de l'analyste une liste de références bibliographiques. Le fichier est alors utilisé comme voie d'accès aux documents, comme instrument de gestion d'une bibliothèque.

Il s'agit là d'un rôle capital par les services qu'il peut rendre aux enquêteurs qui désirent savoir si le témoin qu'ils vont interroger a fait d'autres observations, si le phénomène qu'il rapporte a été observé par d'autres personnes on en d'autres circonstances, ou aux spécialistes de toutes disciplines qui désirent accéder aux documents qui décrivent des phénomènes relevant de leur compétence ou du su jet d'étude qu'ils ont choisi .

Dans tous les cas, on suppose que l'analyste va rechercher les documents et en faire lui-même l'étude. Il n'attend donc pas du fichier une information rigoureuse sur le contenu du document, puisqu'il se propose de l'examiner lui-même, mais simplement la séparation de l'ensemble des documents disponibles en deux lots : celui où il ne trouvera certainement rien à l'intéresser et celui où il a une chance de glaner des informations pertinentes. Il y a donc deux risques à prendre en considération :

- le risque qu'un document intéressant ne figure pas sur la liste après traitement informatique. On souhaite qua ce risque soit nul ;

- le risque qu'un document sans intérêt figure sur la liste. Là, par contre, on est moins exigeant puisque, par hypothèse, le chercheur va opérer son propre tri. En fait, on est d'autant moins exigeant qu' il y a moins de documents à trier - Avec 1000 documents sélectionnés, une probabilité de succès de 0,7 (30 % de documents sans intérêt) peut être gênantes avec 100 documents, on peut descendre à 0,5 et moins.

Les objectifs à atteindre étant ainsi précisés, voyons ce qui en résulte pour le codage.

Un codage complet et scrupuleusement exact des références du document est bien entendu nécessaire. Quant aux variables, destinées à la sélection bibliographique, on notera qu'elles peuvent fort bien être de la forme "il y a de nombreuses données quantitatives dans le rapport", ou "la trajectoire est très bien décrite", sans que ces données quantitatives ou cette trajectoire soient codées par ailleurs.

3.2.3. Fonction statistique

Dans l'approche statistique on s'intéresse moins aux évènements pris individuellement qu'aux propriétés d'ensemble du phénomène OVNI telles qu'elles apparaissent par l'étude de nombreux cas. Quand le nombre des documents à utiliser devient grand, l'analyste ne peut plus utiliser les documents eux-mêmes ; mais doit faire usage des informations codées dans le fichier. Il attend donc du fichier une information rigoureuse sur le contenu du document car, si le codage est mauvais, les statistiques qu'il établira seront nécessairement mauvaises.

Le fichier doit être adapté aux études dès maintenant entreprises ou qui le seront à l'avenir. Pour autant qu'on puisse le savoir à l'heure actuelle, ces études tombent dans l'une des quatre rubriques suivantes :

→ Études spatio-temporelles :
Elles visent à décrire la distribution dans le temps et dans l'espace des observations et à en dégager les propriétés du phénomène qui leur a donné naissance. Elles nécessitent le plus grand nombre de cas possibles, idéalement tous les cas publiés et ce indépendamment de leur qualité ou de leur appartenance à un phénomène OVNI spécifique.

Les variables impliquées au premier chef par cette étude sont : la date, l'heure, la minute, le lieu exact, les conditions d'observation, une description de l'environnement, au moins le type du phénomène observé.

Ces variables sont les plus généralement disponibles sur un cas quelle que soit par ailleurs la qualité du document, elles sont les moins susceptibles d'être déformées. Tous les fichiers informatiques créés à ce jour ont fait une large place à ces données et la presque totalité des travaux scientifiques originaux portant sur le phénomène OVNI, font appel à ces seules données. Cette voie de recherche demeure à l'heure actuelle l'une des plus prometteuses qui soit.

→ Etudes de structure et comportement du phénomène :
Elles visent à décrire les caractéristiques propres des phénomènes observés quant à leurs propriétés lumineuses, cinétiques, géométriques, etc... Elles nécessitent donc le plus grand nombre possible de données par cas, les cas retenus étant ceux qui fournissent les meilleures descriptions du phénomène.

Les variables qu'on souhaite trouver dans le fichier â ce propos sont : les conditions d'observation, la compétence de l'observateur, des descriptions détaillées du phénomène lui-même quant à son apparence et à son comportement.

Certaines de ces variables sont souvent disponibles, mais avec une précision qui peut laisser à désirer : la qualité de l'observation, celle de l'enquête et du rapport qui en résulte influent beaucoup sur la qualité des informations à utiliser. Elles sont aussi les plus susceptibles d'être déformées. Autrement dit, ce type d'étude est important mais difficile. Il y a très peu d'études statistiques de valeur disponibles sur ces problèmes à l'heure actuelle.

→ Etudes psycho-sociologiques :
Elles sont orientées vers les témoins. Quelques données sociologiques et traits psychologiques sont extractibles des documents mais pas de manière systématique. Les rapports donnant des indications nombreuses sont rares. L'étude psychologique directe des rapports existant est donc nécessairement limitée. Cependant, quelques travaux ont été publiés à l'aide des données disponibles ; ils visent contradictoirement, à établir l'originalité ou la banalité des observateurs d'OVNI par rapport à la moyenne de la population.

Par contre, une étude indirecte peut être menée à l'aide des résultats des études spatio-temporelles (aspects sociologiques) ou structuro-comportementales (aspects psychologiques). De tels développements sont dès maintenant bien apparents.

→ Etudes physiques :
Elles sont orientées vers l'interprétation à l'aide de modèles physiques des phénomènes observés ou de leurs effets sur l'environnement. Ce type d'études se heurte aux mêmes difficultés que les études psycho-sociologiques liées à la rareté des informations aptes à les étayer.
Notons que ces études physiques apparaissent comme un développement spécialisé et particulièrement exigeant en bonnes données du point n° 2 : "structure et comportement du phénomène". Précisons aussi pour éviter les malentendus que les études physiques auxquelles il est fait référence ici, sont celles qui visent à confronter les modèles physiques aux observations ou à déduire des grandeurs physiques caractéristiques à partir des observations et non celles qui visent à développer des théories ou des modèles, ce qui est de la physique pure et non pas de la physique appliquée à l'ufologie. L'une se passe très bien du fichier informatique, mais l'autre est contrainte d'y faire appel.

Il est probable que ces études ne se développeront qu'à la suite des enquêtes effectuées spécialement en ce sens par le GEPAN, à partir de cas accidentellement favorables (effets particuliers, radar, traces, ...). Toutefois, d' utiles indications pourraient être extraites des cas publiés, des documents existants, comme le montrent les études de Mc CAMPBELL (15) ou de POHER (5) , si toutefois on pouvait mettre la main sur les documents utilisables. Le but qu'on doit se proposer d'atteindre ici est donc moins de coder en grands détails les indications qu'on peut supposer utiles aux physiciens (à notre sens, c'est impossible à réaliser), qu'à signaler les rapports contenant beaucoup d'informations apparemment utilisables. On développe la fonction bibliographique au détriment de la fonction statistique.

Il est facile de justifier cette option par l'exemple. On peut considérer pour cela l'article de POHER (5) "Proposition d'une méthode permettant de vérifier si les OVNI sont des véhicules". La première méthode qu'il propose consiste "à examiner les variations de la vitesse de l'objet en fonction de son inclinaison sur l'horizontale". Le nombre de cas de prêtant à cette analyse est infime et il est par conséquent hors de question de prévoir un code spécifique pour cette occurence exceptionnelle. La seconds méthode s'applique aux cas faisant état d'un effet mécanique sur l'environnement, à condition de connaître "la pression équivalente à l'effet mécanique constaté et aussi le diamètre de l'objet observé (ou mieux son volume)". Là encore, un collage spécifique serait d' une lourdeur impraticable.

Le problème est de déterminer aussi rapidement que possible les indications potentiellement précieuses pour le physicien de manière à ce que le fichier signale leur présence dans un document, à défaut de pouvoir en réaliser un codage exhaustif. Il convient, dans ce but, d'établir une liste des situations susceptibles de développements physiques. A côté de ces situations particulières qui ne se signaleraient pas spontanément à l'attention des codeurs, il convient de faire les remarques générales suivantes :

- tout modèle physique implique des analyses quantitatives. Il s'ensuit que toutes les données quantitatives ou quantifiable d'un document doivent faire l'objet d'un soin particulier et que la présence de données quantitatives non codées (faute d'une structure horizontale permettant de le faire) doit être notée sans le fichier (le fichier signale ses propres lacunes) ;

- toute description détaillée de trajectoire, de propriétés lumineuses, d'effets sur l'environnement est susceptible de fournir des indications précieuses même si on ne sait pas exactement comment les utiliser aujourd'hui. Là encore, toute présence de détails précis non codés doit être notée dans le fichier.

3.3. CONCLUSION : CLASSEMENT DES VARIABLES

Une liste de variables à utiliser se dégage logiquement de l'étude qui précède. Elle doit viser à traiter la grande majorité des documents et ce dans tous les aspects. Elle doit s'attacher en outre à satisfaire le plus grand nombre d'applications possibles, tant par le mode d'utilisation (sélection bibliographique, calculs statistiques), que par les questions posées (physique, sociologie, etc...).

Il n'est pas inutile avant de poursuivre de fournir, non pas une liste complète des variables, mais un classement de celles-ci, à l'aide du contenu descriptif qu'elles fournissent.

Le classement suivant nous semble particulièrement utile à la fois conceptuellement et pratiquement :

- les variables documentaires :
  elles décrivent le document en tant que tel : elles fournissent les références bibliographiques complètes, indiquant s'il s'agit d'un document primaire ou secondaire, et tentent de préciser ses qualités et défauts ;

- les variables observationnelles :
  elles décrivent tout ce qui concerne l'observation et les observateurs c'est à dire le contenu du document. On peut les subdiviser elles-mêmes en deux catégories :

  - les variables extrinsèques :
    ce sont les variables observationnelles relatives aux conditions d'observation au sens large (localisation spatio-temporelle, description de l'environnement, du ou des témoin(s))

  - les variables intrinsèques :
    ce sont les variables observationnelles relatives au phénomène allégué : localisation dans l'environnement, forme, structure et dimensions, effets lumineux, sonores, odorants, physiques et biologiques, trajectographie.

Un classement plus précis en sept catégories ou séquences peut également rendre service :
- séquence 1 : description du document,
- séquence 2 : localisation spatio-temporelle de l'évènement,
- séquence 3 : environnement : météo, visibilité, géographie,
- séquence 4 : témoins, véhicules, méthode d'observation, crédibilité
- séquence 5 : phénomène : éléments quantitatifs,
- séquence 6 : phénomène : éléments qualitatifs et type de phénomène,
- séquence 7 : intérêt de l'évènement, explications possibles, etc...

4. PROBLEMES DES MODALITES ("COMMENT CODER" ?)

Les problèmes liés à la structure d'ensemble du fichier ayant été discutés, Il reste à étudier la structure des variables. En effet, chaque variable est divisée en modalités soit par exemple la variable "méthode d'observation", les catégories prévues à l'avance : "oeil nu", "Jumelles", "longue-vue", etc... constituent les modalités de cette variable. Nous abordons ainsi le problème des techniques de codage . Nous envisageons d' abord les problèmes posés par la création des modalités (§4.1) puis ceux posés en pratique par le codage des documents (§9.2).

4. 1. CODAGE D' UN POINT DE VUE THÉORIQUE

Nous avons vu que la variable est l'unité de construction de la structure horizontale du fichier. Une telle variable peut être considérée comme une question posée à laquelle il convient de répondre en faisant usage des informations contenues, dans le document. On peut convenir d'appeler ces informations éléments ou items à coder. L'élément est l'ensemble de mots, d'expressions et de phrases utilisés dans le document qui permet de coder une variable.

4.1.1. Propriétés des éléments : dimensionnalité

La propriété la plus importante d'un élément est sa dimension. Un élément est unidimensionnel lorsqu'il ne peut prendre qu'un seul état parmi plusieurs états discrets (sex. : sexe du témoin, nom de la commune où a eu lieu l'observation alléguées) ou une infinité d'états continus localisés sur un axe unique (ex. : dimension, distance, altitude, durée). Dans ce cas, une variable unique à variation continue ou discontinue suffit à décrire l'élément. Dans tous les autres cas, l'élément est multidimensionnelle et il faut faire usage pour l'expliciter d'un grand nombre de paramètres.

Considérons par exemple, un objet quelconque. S'il s'agit de décrire sa forme géométrique, on voit qu'un grand nombre de paramètres est nécessaire pour peu qu'il soit irrégulier ou complexe : la forma est une caractéristique multidimensionnelle par excellence et c'est ce qui rend si difficile l'étude automatique des images et la reconnaissance automatique des formes. Par contre, la masse du même objet est un paramètre unidimensionnel ; un seul nombre suffit pour épuiser le sujet. Remarquons qu'il existe des caractéristiques intermédiaires : la dimension de l'objet par exemple. S'il s'agit d'une boîte, il y a trois dimensions : longueur, largeur, hauteur, sans ambiguité. S'il s'agit d'une chaise, on peut convenir de considérer les dimensions hors-tout. On peut même convenir dans les deux cas de ne retenir que la plus grande dimension.

Ceci introduit une nouvelle idée qui est à la base même de toute opération de codage à savoir la réduction à l'unidimensionnalité d'informations multidimensionnelles. Le nombre total de variables est limité, et chaque variable est multidimensionnelle au sens de la définition précédente : la réduction s'accompagne donc nécessairement d'une perte d'information. Il faut faire en sorte que l'information originelle conserve suffisamment de relations avec la ou les caractéristique(s) sensée(s) la représenter pour avoir la valeur des schémas et des caricatures qui déforment certes, mais dans le sens de l'essentiel. Pour être faite de manière reproductible, une réduction doit faire usage de commentaires explicites de règles de transformation à consigner dans un livret de codage (voir ci-après). Une information multidimensionnelle est aisément réductible s'il est possible de définir des règles de transformation simples et générales, et si l'on dispose d'un nombre suffisant de variables. Soit, par exemple, à coder les conditions météorologiques. Si on ne dispose que d'une seule variable, la tâche de transformation est ardue et nécessairement déformante, avec plusieurs variables, elle devient beaucoup plus facile.

La multidimensionnalité n'est pas l'exception mais la règle. Toute variable ou presque doit donc être accompagnée des règles qui permettent d'assurer la réduction conventionnelle de l'élèment correspondant. Deux cas sont à distinguer :

- on peut, dès à présent, proposer des règles de réduction satisfaisantes pour tout ce qui dans un document est numérique (nombre de témoins, d'objets, ...), dimensionnel (taille de l'objet, altitude, distance, ...), ou temporel (durée du phénomène,...). Des règles de transformation sont nécessaires même pour ce qui parait de prime abord unidimensionnel comme les nombres et les durées. Il se peut par exemple, que les témoins voient initialement trois objets en formation qui disparaissent à leur vue, puis un objet solitaire. L'élément "nombre d'objets" est multidimensionnel. On peut décider par convention de coder "4 objets vus", mais il est clair que cette réduction entraîne une perte d'information. A lire le code, on pourrait croire qu'ils ont vu 4 objets simultanément, or il n'en est rien. La réduction à opérer s'impose lorsque l'une des dimensions de l'élément, un de ses aspects est plus important (témoin principal , dimension maximum, altitude minimum,, etc...)

- dans d'autres cas, tels que les descriptions du lieu, du témoin, du phénomène, la réduction est plus difficile. Dans une description de trajectoire, de forme, de jeux de lumière, les dimensions caractéristiques ne s'imposent pas. La définition d'états discrets imposée par le codage, est nécessairement vague et arbitraire : la précision du codage s'en ressent. C'est ici que les mises au point seront les plus longues avant d'obtenir des règles de transformation et de codage adéquates : les éléments qualitatifs nécessitent des études préliminaires.

De toute manière, un procédé de réduction conservera toujours quelque arbitraire. Aussi, faut-il savoir qu'un élément aurait pu être codé différemment : aucun système de collage ne peut épuiser le contenu d'un document.

4.1.2. Propriétés des variables : échelle de mesure

Lorsqu'on a réussi une réduction à l'unidimensionnalité, il reste à effectuer une "mesure" au long de cet axe, à y définir une échelle quelconque. On distingue 4 types de "mesures" ou échelles qui sort des plus faibles au plus fortes : les échelles nominales, ordinales, d'intervalles et de rapports. Les deux premières s'appliquent aux caractères qualitatifs, les deux dernières aux caractères quantitatifs (7) (8).

L'échelle nominale s'applique aux caractères qui ne peuvent être que classés à l'aide d'un critère précis, par exemple : sexe, état civil d'un témoin, forme d'un objet, ... La seule propriété formelle de cette échelle est d'être structurée par une relation d'équivalence (=) : les membres d'une classe sont équivalents vis à vis de la proprété utilise pour construire l'échelle. Les seules statistiques descriptives admissibles sont fondées sur un calcul de fréquence et les tests non paramétriques correspondants (X^2, notamment).

L'échelle ordinale s'applique aux caractères qui peuvent être classés et dont les classes d'équivalence peuvent en outre, être ordonnées en faisant usage d'une relation du genre "est plus grand que", "est préféré à", etc... Toutefois, les intervalles entre ces classes ne sont pas définis. Elle incorpore donc une relation d'équivalence (=) et une relation d'ordre (>). Les statistiques applicables à l'échelle nominale s'appliquent encore mais s'y ajoutent les "tests de rang".

L'échelle d'intervalles est une échelle ordinale où les distances entre classes sont connues. Cette échelle est caractérisée par une unité de mesure constante mais cette unité et le point origine sont

arbitraires. L'exemple classique est celui de la mesure des températures sur les échelles d'intervalle CELSIUS ou FAHRENHEIT. Peu au pas d'applications ufologiques.

L'échelle de rapports est la plus forte, elle a toutes les propriétés d'une échelle d'intervalles et en plus a une origine fixée. Seule l'unité de mesure est arbitraire. Elle se définit par 4 propriétés : équivalence, ordre, rapport à intervalles, rapport de points. Elle s'applique à l'ufologie dès lors qu'il y a évaluation précise de distances, d'angles, de durées, etc...

On notera que les variables issues d'éléments unidimensionnels, peuvent se mesurer suivant les cas dans l'une ou l'autre des 3 échelles principales mais que celles issues d'éléments multidimensionnels ne peuvent être mesurées que dans des échelles ordinales ou nominales. On a bien entendu toujours intérêt à utiliser l'échelle la plus forte possible et il y a progrès lorsqu'on détermine des critères permettant de transformer une variable jusque là nominale en variable ordinale. Ainsi, une variable aussi multidimensionnelle que le type de phénomène selon SAUNDERS, qui pourrait sembler purement nominale, peut être considérée comme une mesure ordinale de la force de l'interaction phénomène/témoin.

L'hétérogénéité des documents invite à prendre les précautions suivantes :

1) On prévoiera pour chaque variable, autant d'échelles différentes que possible, pour permettre le choix correspondant le mieux à la précision disponible dans le rapport. Ainsi, les modalités relatives à la variable "distance minimum phénomène/témoin" devront rendre possible un collage numérique (échelle de rapport) ou l'affectation à une classe de distance (échelle ordinale) ou même une simple qualification de cette distance du genre "près", "loin" (échelle quasi nominale) ;

2) Pour les échelles ordinales et nominales, plusieurs découpages en classes de précision croissante pourront être prévus. Il convient toutefois, de définir ces classes de manière à ce que la transcodabilité interne, c'est à dire le passage d'une classe précise à une classe moins précise, reste possible. Le fait qu'un gain en effectif puisse être obtenu au prix dune perte en information est une propriété très utile dans bon hombre d'études statistiques

4.2. CODAGE D'UN POINT DE VUE PRATIQUE

La qualité du codage effectué est limitée par les imperfections du système de codage d'une part et par l'équation personnelle du codeur d'autre part.
Examinons ces deux points.

4.2.1. Limitations liées au système de codage - Livret de codage

Lorsqu'un code est élaboré un peu a priori sans une connaissance approfondie des documents, on constate généralement que son utilisation est malaisée : dans nombre de cas, il ne s'applique pas aux informations effectivement présentes, et lorsqu'il s'applique, de nombreuses ambiguités se manifestent. Une telle situation est très dommageable parce que le temps gagné sans l'élaboration du code, est perdu et bien au-delà dans son utilisation en routine et que la qualité du codage obtenu est médiocre. La co-adaptation des documents à coder et du système de codage ne peut être atteinte que par un processus de perfectionnement par essais et erreurs : il faut absolument que les difficultés soient transférées dans : la mise au point du système plutôt que dans son utilisation.

Mais, il ne suffit pas qu'un système de codage ait été l'objet dune mise au point soignée, encore faut-il qu'il soit accompagné d'un livret de codage (6). Un tel livret ne se contente pas de donner

une liste de variables et de leurs modalités mais il en explicite dans chaque cas le mode d'emploi à l'aide de définitions, d'exemples, de contre-exemples (erreurs à ne pas commettre), etc...
Il donne en particulier, les règles de réduction conventionnelles pour les éléments multidimensionnelle.

Un chercheur isolé peut dans une certaine mesure se passer d'un livret de codage parce qu'il est à même d'utiliser de manière cohérente un code même s'il n'a pas fait l'effort de l'expliciter. Encore faut-il qu'il travaille régulièrement et rapidement sinon une dérive se produit au cours du temps par une réflexion plus ou moins inconsciente, des adaptations qui changent imperceptiblement mais sûrement le sens des codes employés. Un tel mécanisme a des conséquences encore plus funestes lorsque le codage est réalisé par une équipe car chaque codeur élabore sa propre définition et évolue à sa façon. Dès lots, le code utilisé ne recouvre plus une réalité constante, il devient irreproductible. Or, il est capital pour l'utilisation du fichier, que le sens de chaque code soit clair et unique car on ne saurait progresser sur un fond d'à peu près.

Il ne faut pas se leurrer : un livret de codage bien au point est nécessairement d'une certaine complexité. Il doit exiger du codeur inexpérimenté un effort de compréhension. Cet effort est le prix qu'il faut payer pour obtenir un codage reproductible et au bout du compte facile parce que dépourvu d'ambiguité . En l'absence de ces Indications qui ralentissent son travail au début, le codeur peut croire la tâche plus facile, mais confronté à la première difficulté venue, il hésite, choisit un peu au hasard et il en résulte une perte importante de qualité : à la limite l'indication codée se révèle dépendre autant de l'équation personnelle du codeur, du hasard du moment que du document lui-même. Une telle situation est bien entendu inacceptable et il faut tout mettre en oeuvre pour y remédier.

Un livret de codage n'est jamais achevé, même en supposant le système de codage constant, aussi faut-il prévoir des procédures pour en assurer l'amélioration constante :

- une première procédure pourrait reposer sur les codeurs eux-mêmes qui, lorsqu'ils seraient confrontés à des problèmes difficiles à résoudre, à l'aide des indications du livret, rempliraient un formulaire spécial. Ceci suppose des codeurs déjà bien expérimentés ayant intégré les règles en application à un instant donné de manière à faire la part des difficultés personnelles d'apprentissage et des difficultés liées aux insuffisances du livret :

- une seconde procédure consisterait à tester régulièrement les codages en comparant les indications de deux codeurs expérimentés sur un même document. Les inévitables fautes d'attention mises à part, les divergences signaleraient des ambiguités à supprimer dans les règles en cours.

La rédaction d'un tel livret et sa mise à jour permanents auraient l'avantage d'obliger à une explication constante des règles de codage suivies, sans lesquelles il est impossible d'accéder à une quelconque objectivité dans le codage. Il permettrait également de former rapidement les nouveaux codeurs de manière rationnelle, et non "sur le tas", ce qui éviterait le retour indéfini aux mêmes errements.

4.2.2. L'équation Personnelle du codeur

La transformation du document en modalités codées exige la lecture, la compréhension et, dans une certaine mesure, l'interprétation du texte, par un codeur dont la personnalité va intervenir à chacune de ces étapes. Les risques d'interprétation subjective sont donc réels, notre expérience personnelle nous conduit a en distinguer deux principaux :

1) Le premier risque est de "ne pas chercher à bien faire". Il résulte de la frustration qu'engendrent souvent certains documents ou certaines observations. Le codeur peut avoir l'impression que l'imprécision inhérente à l'observation ou au document, ne justifie pas de sa part un effort de compréhension ou d'analyse. Le même sentiment peut être partagé par l'analyste lorsqu'il aborde la création de certains codes : "à quoi bon expliciter ces concepts puisque les données dont nous disposons sont souvent mal recueillies !". Cette tentation d'approcher les documents de manière globale et approximative est trop fréquente pour ne pas être signalée ici comme un danger majeur dont il convient de se prémunir avant tout. Si les codes sont mal définis et si le codeur les utilise avec légèreté, l'imprécision inhérente au collage devient telle que sa valeur scientifique disparaît, le résultat est bon à mettre au panier. L'imprécision des documents exige beaucoup de rigueur et de précision au niveau du système de codage et de la pratique du codage. Le paradoxe central du codage est : à document imprécis, codage précis.

2) Le second risque provient, au contraire, du désir de "trop bien faire" en dépassant le contenu objectif du document. C'est un fait psychologique que la lecture d'un document suscite spontanément dans l'esprit une image intuitive de l'évènement qui tend à se substituer à ce qui est explicitement écrit. Certaines lacunes du document ne sont pas perçues et l'imagination suppléait ces manques. Sur cette tendance naturelle, peut en outre se greffer une rationalisation excessive, si bien qu'en fin de compte ce n'est pas le document qui est codé mais la reconstitution opérée par le codeur.

Il ne faut pas cependant en conclure que toute interprétation est préjudiciable. En effet, le concept de "contenu objectif" du document n'est pas clair. Il ne s'agit pas d'en tester dans tous les cas à la lettre du document mais d'accéder à son sens. Ainsi, convient-il pour coder une variable de prendre simultanément en compte tous les passages du document qui la concernent et d'en effectuer la synthèse, ce qui implique une certaine reconstitution. Mais il faut procéder avec beaucoup de prudence de manière à ce que ces interprétations éclairent le document plus qu'elles ne le déforment, et entre les deux la voie est étroite. On aura garde d'oublier que le fichier n'est pas relatif à une observation mais à un document.

Il est beaucoup plus fructueux de discuter ces risques d'interprétation subjective variable par variable qu'*in abstracto* comme nous venons de le faire. Toutes les variables ne présentent pas les mêmes dangers. En outre, pour aider le codeur à l'objectivité, des variables distinctes peuvent être créées pour un même élément : l'une de contenu (ce que dit le document), l'autre de commentaire (ce qu'en pause le codeur). Dans tous les cas, le livret de collage doit expliciter la méthode à suivre dans les situations délicates. Finalement, toute la méthodologie propre à la réalisation d'un fichier vise spécifiquement à réduire ces risques de subjectivité. Toutefois, rien ne peut remplacer la formation et la rigueur méthodologique des codeurs eux-mêmes, et la création d'une ambiance favorable à l'acquisition de cette expérience et à l'exercice de cette rigueur.

5. LES FICHIERS EXISTANTS

Pour définir le fichier, il convient de tenir compte non seulement de la nature des documents disponibles, des études à réaliser et des impératifs du codage, mais aussi des fichiers réalisés antérieurement par d'autres chercheurs.

5.1. DESCRIPTION SOMMAIRE

1) Le premier chercheur en date à s'être attelé à la réalisation d'un fichier informatique est Jacques VALLEE (9) (10) au début des années 60. Il a ainsi réalisé un catalogue mondial de

3 073 cas. L'unité de base du fichier est la manifestation du phénomène et non le document. La localisation spatio-temporelle y est codée avec un soin tout particulier. Les autres variables extrinsèques sont codées de manière sommaire dans le but de repérer les cas obéissant à un certain nombre de critères pré-sélectionnés. Le codage des variables intrinsèques est intéressant mais non systématique.

2) Le second chercheur par ordre chronologique est David SAUNDERS (4) à partir de 1967. Son fichier, appelé UFOCAT, a été réalisé initialement dans le cadre du Comité CONDON et à titre privé par la suite. Il est actuellement développé au sein du Center for UFO Studies l'Organisation privée, dirigée par le Dr. HYNEK, EVANSTON, Illinois). Il s'est accru à raison de 10 000 entrées par an, il comporte donc aujourd'hui plus de 100 000 entrées. L'unité de base y est le document. Ce fichier international est de loin la plus importante collection de données ufologiques actuellement disponibles sous forme informatisée bien que tous les documents fichés n'aient pas été codés complètement. D. SAUNDERS a développé un certain nombre des idées ébauchées par VALLEE. Il conserve toute son importance à la localisation spatio-temporelle et même la développe. Il Introduit un codage assez complet des caractéristiques de la source utilisée (variables documentaires). Il utilise une typologie du phénomène fort intéressante, différente de celle de VALLEE et de celle introduite plus tard par HYNEK (13) .

3) Claude POHER (11) a développé son fichier de manière complètement indépendante pour autant qu'on puisse en juger par la structure horizontale de celui-ci. La prééminence accordée à la localisation spatio-temporelle disparaît et, en conséquence, la place accordée aux variables intrinsèques s'accroît : le codage en est plus systèmatique que chez les deux autres précédents. L'unité de base est encore la manifestation du phénomène. Ce fichier compte 736 entrées.

4) Le fichier actuellement développé par le GEPAN (6) (12) , à titre expérimental, sur la base des rapports de Gendarmerie, dérive directement de celui de POHER et est totalement indépendant de ceux de VALLEE et de SAUNDERS. Certaines variables supplémentaires apparaissent par rapport à POHER (département, témoins potentiels, âges, sexes, catégories socioprofessionnelles des témoins, direction azimutale, type selon HYNEK) mais d'autres disparaissent en raison de leurs faibles fréquences (lumières ponctuelles, odeurs, effets physiques, biologiques ou psychologiques, description et même mention d'occupants).

5) Les fichiers précédents peuvent être tour qualifiés de "généralistes" en ce sens qu'ils portent sur tour les types de phénomènes observés, incluent tous les types de variables observationnelles (même s'ils insistent plus sur les variables extrinsèques ou intrinsèques suivant les cas) et qu'ils visent à traiter plutôt sommairement le plus grand nombre possible de cas.

Nous avons été amenés à développer un fichier (18) à partir de 1973, obéissant à une philosophie radicalement différente, que l'on peut qualifier de "spécialiste" par opposition aux précédentes dont les caractéristiques sont les suivantes :

- il porte sur un type de phénomène unique et très spécifique les atterrissages avec exhibition d'occupants ;

- il est orienté vers la description de la structure et du comportement du phénomène et une explicitation des processus d'observation, de divulgation et d'enquête ;

- l'effort a porté non sur l'extension du nombre de cas (stabilisé provisoirement à 70) mais sur l'extension (au cours du temps) du nombre de variables de manière à opérer une description exhaustive des cas ;

- la méthodologie du codage a été l'objet d'une attention particulière quant à la définition et à l'application des modalités. Le procédé des "classes disjonctives complètes" a été privilégié.

Cette approche est mentionnée ici pour insister sur le fait qu'un fichier "généraliste" n'est pas adapté à toutes les études et que des fichiers "spécialistes" devront être créés dans son orbite, si possible par les chercheurs eux-mêmes .

5.2. COMPARAISON DES FICHIERS GENERALISTES

Pour comparer les fichiers, nous avons dressé une liste de 60 variables en raison de leur présence dans l'un ou l'autre des systèmes de codage ou de l'intérêt qu'elles nous semblaient avoir. Nous avons noté pour chaque fichier les variables de cette liste qui sont codées et celles qui ne le sont pas (tableau 1). Nous constatons ainsi qu'il n'y a que 6 variables seulement qui, disposent d'un code spécifique (c'est à dire ne sont pas mêlées à d'autres variables) dans les quatre fichiers. Ce sont :

- date
- heure/minute
- nombre témoins
- nombre d'objets
- dimension de l'objet
- latitude/longitude

Noter que toutes ces variables sont quantitatives et que la moitié d'entre elles concernent la localisation de l'événement. Si on retient également les variables qui ont un codage spécifique dans 3 fichiers ou spécifique (plusieurs variables logiquement distinctes confondues en une seule) dans les 4, il vient :

- âge des témoins
- durée d'observation
- forme
- bruit
- type du phénomène
- localisation/méthode d'obs. des témoins
- trajectoire
- luminosité
- effets physiques
- crédibilité/étrangeté

Pour mieux apprécier les pôles d'intérêt de chaque auteur, on peut regrouper les 60 variables en 7 séquences (voir § 3.3) chacune décrivant un des aspects du document ou de l'observation. Il suffit alors de se faire une idée de l'importance de chacune des séquences, de compter le nombre de colonnes de carte perforée qui leur est consacré (tableau 2).

DOCUMENT	V	P	G	S
1 Nature	o	●	●	●
Identification	o	●	o	●
Réf. intra-source	o	o	o	●
Source indirecte	o	o	o	●
Transformations	o	o	o	o
Longueur	o	o	o	o
Inf. graphiques	o	o	o	o
VARIABLES EXTRINSEQUES				
2 Date	●	●	●	●
Heure/minute	●	●	●	●
Temps sidéral	o	o	o	●
Nom du lieu	⧱	o	o	●
Dép./province		o	●	●
Pays/Etat	●	o	o	●
Lat./longitude	●	●	●	●
Position orthoténique	o	o	o	✱
3 Etat de surface	o	o	o	⧱
Précipitations	o	⧱	⧱	⧱
Couverture nuageuse	o	⧱	⧱	●
Orage	o	⧱	⧱	⧱
Visibilité	●	⧱	⧱	⧱
Témoins potentiels	o	o	●	o
Particularités locales	o	o	o	●
4 Nombre de témoins	●	●	●	●
Nom des témoins	o	◉	o	●
Age	o	●	●	●
Sexe	o	o	●	●
Catégorie socio-prof.	⧱	●	●	✱
Localisation	◉	⧱	⧱	●
Méthode d'obs		●	●	⧱
Photo/film	◉	⧱	⧱	◉
Témoins indép	o	o	o	⧱
Témoins récurrents	o	o	o	o

VARIABLES INTRINSEQUES UNIDIMENSIONNELLES	V	P	G	S
5 Nombre d'objets	●	●	●	●
Durée d'obs	o	●	●	●
Dimension	●	●	●	●
Altitude	o	●	●	o
Distance	o	●	●	o
Direction azimutale	o	o	●	o
VARIABLES INTRINSEQUES MULTIDIMENSIONNELLES				
6 Trajectoire	⧱	●	●	◉
Vitesse	◉	●	●	o
Accélération	⧱	●	●	o
Formations	⧱	o	o	✱
Forme	◉	●	●	◉
Structure	⧱	o	o	o
Luminosité	⧱	●	●	◉
Couleur	◉	●	●	o
Lumières ponctuelles	o	●	o	o
Etat de surface	o	o	o	o
Bruit	o	●	●	◉
Odeur	◉	●	o	o
Effets physiques	●	●	o	◉
Effets biologiques	o	●	o	◉
Occupants	o	✱	o	◉
7 Type	●	o	●	●
Sous-type	o	o	o	✱
8 Explication selon source	o	o	o	●
Type d'explication	o	o	o	●
Crédibilité témoins	⧱	●	●	✱
Etrangeté phénomène	◉	●	⧱	✱
Intérêt document	⧱	o	◉	✱

LEGENDE : V - fichier VALLEE (61-66) ; P - fichier POHER (71)
G - fichier GEPAN (78-79) ; S - fichier SAUNDERS (67-77)

1 - document ; 2 - Localisation ; 3 - Environnement ; 4 - témoins
5 - phénomène "quantitatif" ; 6 - phénomène "qualitatif" ;
7 - typologie ; 8 - intérêt.

✱ codé avec des variables ou modalités nombreuses ; ● codé spécifiquement
o non codé ; ⧱ codé simultanément ; ✱ prévu mais non décrit par SAUNDERS (4)
◉ codé de manière non systématique ou sommaire ou à l'aide de modalités de la variable "sous-type" (chez SAUNDERS).

- TABLEAU N° 2 -

SEQUENCES	SAUNDERS (4)		POHER (11)		GEPAN 4 (12)	
1 - Document	21	15 %	3	4 %	2	3 %
2 - Localisation	64	46 %	15	20 %	23	34 %
3 - Environnement	3	2 %	2	3 %	2	3 %
4 - Témoins	24	17 %	5	7 %	8	12 %
5 - Phénomène quantitatif	9	6 %	9	12 %	17	25 %
6 - Phénomène qualitatif	5	4 %	40	53 %	13	19 %
7 - Intérêt	14	10 %	2	3 %	2	3 %
	140		**76**		**67**	

On voit que SAUNDERS consacre près de la moitié de ses colonnes disponibles à localiser le phénomène tandis que POHER en consacre plus de la moitié à le décrire. Le système GEPAN 4ème version, équilibre mieux les deux aspects mais au bénéfice de la description intrinsèque.

Il est donc clair que des options extrêmement divergentes peuvent être prises quant à la structure horizontale d'un fichier, qui ne dépendent ni des documents disponibles (tous les auteurs ont les même documents ou des documents comparables en main) ni des moyens disponibles (tous les auteurs ont été aidés de collaborateurs) mais principalement des pôles d'intérêt scientifiques, des études que ces auteurs se sont proposés de réaliser pour mieux comprendre le phénomène OVNI (VALLEE et SAUNDERS se sont intéressés à la théorie de l'orthoténie due à Aimé MICHEL (17) et au problème des "vagues" d'où leur insistance sur les aspects spatio-temporels).

5.3. IMPORTANCE D'UFOCAT

L'existence du fichier UFOCAT de SAUNDERS est extrêmement positive pour le GEPAN par l'expérience mise à notre disposition dans "The UFOCAT Code Book" et par la possibilité d'opérer des échanges à l'avenir aux conditions stipulées par cet auteur. Mais il constitue un défi :

- d'un point de vue quantitatif, par l'avance déjà prise et par la vitesse d'accroissement du fichier,

- d'un point de vue qualitatif, venant après SAUNDERS , il faut autant que possible utiliser des méthodes de codage supérieures aux siennes.

Dès que la structure horizontale du fichier sera déterminée, il conviendra de mettre au point un algorithme de transcodage destiné à convertir les codes GEPAN en codes UFOCAT.

6. CONCLUSIONS

6.1. DEFINITION DU FICHIER

Le fichier informatique général à développer par le GEPAN, tel qu'il résulte de l'analyse précédente, possède les caractéristiques suivantes :

- il concerne l'ensemble des observations faites sur le territoire français. Le mot "observation" est entendu dans son sens le plus large, à savoir : l'observation alléguée par un ou plusieurs témoins, d'un phénomène quelconque qu'ils ont eux-mêmes attribué à un OVNI ou que d'autres ont ainsi Interprété. Cette définition est indépendante de la véracité des témoins, de la qualité de leur déposition de celle de l'enquête qui en ait résultée ou de celle

du document qui fait état de l'observation alléguée ; elle est indépendante de la possibilité d'expliquer ou non cette observation de manière conventionnelle. Le seuil d'acceptation d'une observation pour le fichier est donc placé au plus bas pour assurer la généralité de celui-ci et la possibilité d'y définir de nombreux sous-ensembles pour des études particulières allant de la sociologie à la physique des phénomènes allégués ;

- l'unité de construction du fichier n'est pas l'observation mais le document faisant état dune observation. Tout document faisant état d'une observation au sens précédent peut être inclus dans le fichier. Cette option, outre qu'elle préserve une utilisation sociologique ou historique par exemple, est la seule qui puisse pratiquement assurer à terme l'exhaustivité du fichier : pour les évènements anciens, les documents de seconde main peuvent seuls subsister, ne pas les coder c'est perdre toute trace de certaines observations ; pour les évènements récents les documents de seconde main signalent l'existence d'évènements dont il convient de rechercher activement les sources primaires ;

- le codage utilisé doit contenir suffisamment de variables pour permettre une description correcte des documents moyens voire légèrement supérieurs à la moyenne. Il s'ensuit par exemple, que nombre d'aspects des rencontres rapprochées ne pourront être codés faute de variables disponibles pour en rendre compte. Les variables disponibles doivent être telles qu'elles permettent le traitent du fichier suivant deux axes complémentaires. Le premier axe vise un simple traitement bibliographique : le fichier est utilisé pour signaler les observations ou documents remarquables étant entendu que les informations recherchées sont présentes dans les documents originaux mais pas nécessairement dans le fichier informatique. Le second axe vise au traitement statistique des informations effectivement présentes dans le fichier sans qu'il soit jamais nécessaire de recourir aux documents originaux que le fichier remplace complètement. Cette double possibilité d'utilisation garantit autant que faire se peut les intérêts de tous les chercheurs qui dès maintenant ou à l'avenir auront besoin d'un fichier informatique pour trier la masses inextricable autrement, des documents ufologiques disponibles, même et surtout en l'absence des codes explicites qui leur auraient été nécessaires.

Le fichier peut donc être qualifié de <u>général</u> à l'aide des critères suivants :

- il s'intéresse è tout évènement (observations, traces, etc...) d'intérêt ufologique quelconque ;

- il utilise tous les documents ufologiques disponibles ;

- il code toutes les variables dont la fréquence d'occurence dans les documents disponibles est suffisamment élevée et néglige sciemment les autres ;

- il vise à rendre service à tous les utilisateurs potentiels d'un fichier informatique, présents et à venir, dans la mesure du possible.

6.2. STRUCTURE HIERARCHIQUE DU FICHIER

Il semble préférable de ne pas concevoir le fichier comme un bloc monolithique où toutes les entrées de toutes les variables auraient a priori la même importance. Les entrées sont hiérarchisées : une entrée primaire est plus importante qu'une entrée secondaire, une entrée primaire bien décrite qu'une sommairement décrite, etc... Il en va de même des variables comme nous allons le voir, les variables documentaires sont plus importantes que les variables observationnelles, et parmi ces dernières les extrinsèques plus que les intrinsèques.

Ces priorités eu égard aux documents et aux variables sont importantes car elles permettent de développer le fichier par étapes et de s'attaquer à la mise au point des techniques de codage d'autant plus sophistiquées qu'on ne vise pas à les appliquer à tous les documents de manière différenciée.

Il n'est pas inutile, pour fixer les idées, de décrire en la justifiant l'organisation hiérarchique des variables, car elle découle de la prise en compte simultanée de nombre d'aspects discutés dans ce travail. On peut regrouper de ce point de vue les variables en cinq sous-ensembles qui sont par ordre d'importance décroissante :

1) Les variables documentaires :
qui fournissent des renseignements sur le document, en premier lieu ses références bibliographiques, forment la plus importante séquence de variables parce que pour coder un document, il faut d'abord savoir qu'il existe et où il se trouve. Elles permettent d'ouvrir une entrée pour un codage ultérieur s'il y a lieu ;

2) Les variables extrinsèques :
codant les coordonnées spatio-temporelles de l'évènement et le nom du témoin viennent en second lieu car :

- elles permettent d'ordonner logiquement le fichier en rangeant les observations par ordre chronologique et en affectant chaque entrée au bloc qui est le sien,

- les variables concernées sont faciles à déterminer et ne posent pas de problème technique pour leur codage.

Si le document codé est secondaire, il n'est pas urgent de coder au-delà à condition que le document primaire correspondant soit déjà codé ou qu'on ait bon espoir d'en disposer rapidement.

3) Les variables extrinsèques codant les données relatives aux témoins et à l'environnement.
Ainsi que la ou les variable(s) codant le type du phénomène observé suivent logiquement car elles constituent en conjonction avec les variables (2), la matière principale des études spatio-temporelles.
Celles-ci nécessitent par leur nature même le plus grand nombre possible de cas quels qu'ils soient. Ces variables peuvent cependant poser des problèmes techniques.

4) Les variables intrinsèques quantitatives codant les aspects quantitatifs du phénomène (nombre d'objets, dimensions, etc...) permettent d'aborder les études de structure et de comportement. De toutes les variables concernant directement le phénomène, ce sont les plus propres à un codage précis du fait de leur réduction aisée à l'unidimensionnalité. Ce sont par ailleurs celles qui se prêtent le mieux à l'analyse scientifique. Elles doivent donc faire l'objet d'une attention particulière. De plus, on est dès maintenant en mesure de proposer pour elles de bonnes règles de codage dont il est peu probable qu'elles soient remises en cause.

5) Les variables intrinsèques qualitatives codant les aspects qualitatifs du phénomène sont de toutes les plus difficiles à coder et les plus sujettes à être déformées par des appréciations subjectives. Ces aspects sont généralement difficiles à ramener à une caractéristique unidimensionnelle soit en raison de leur complexité propre (forme et structure d'un objet, etc...) soit de leur évolution dans le temps (par ex. : comportement de l'objet sur sa trajectoire faisant intervenir des variations de vitesse, d'assiette, de rayonnement, etc...). Il semble préférable de ne pas viser d'emblée une solution définitive mais d'améliorer progressivement les codes en fonction de l'avancement de nos connaissances car, comme le remarque très justement A. ESTERLE (6) :

"*le caractère remarquable ou significatif de tel ou tel critère ne peut être lui-même que le résultat d'une analyse statistique préalable"*. En tout cas, les insuffisances du codage de contenu peuvent dans une certaine mesure être suppléées par un codage du genre "indications bibliographiques".

Cette conception du fichier général se prolonge logiquement par le développement parallèle de fichiers spécialisés. Ces fichiers peuvent être définis sur des sous-ensembles du fichier général et sont destinés à l'exécution de recherches particulières qui nécessitent des codages très difficiles (et alors le chercheur intéressé s'en charge lui-même) ou qui ne s'appliquent qu'à un nombre restreint de documents (voir § 3.1 et § 5.1. - 5°).

REFERENCES

(1) VALLEE J, in BALLESTER OLMOS VJ (1976) - A catalogue of 200 type-I UFO events in Spain and Portugal. CUFOS
(2) Pour tous ces problèmes méthodologiques voir FOURASTIE J. (1966), Les conditions de l'esprit scientifique - Gallimard, collection "idées"
(3) ROSPARS JP, (1978) - Réflexions sur l'emploi des statistiques dans l'étude du phénomène OVNI - GEPAN (juin 78, tome 5, annexe 22, pages 5 à 12).
(4) SAUNDERS D, (1977) - The UFOCAT Code Book - CUFOS
(5) POHER C, (1978) - Proposition dune méthode permettant de vérifier si les OVNI sont des véhicules - GEPAN, (juin 78, tome 5, annexe 15).
(6) ESTERLE A, (1978) - Préparation d'un fichier national d'observations, GEPAN, (juin 78, tome 5, annexe 21).
(7) PINTY JJ et GAULTIER C (1971) - Dictionnaire pratique de mathématiques et statistiques en sciences humaines - Editions universitaires.
(8) SIEGEL S (1956) - Non parametric statistics for the behavioral sciences, Mc GRAW-HILL Kogakusha.
(9) VALLEE J, (1965) - Anatomy of a phenomenon - REGNERY.
(10) VALLEE J, (1966) - Les phénomènes insolites de l'espace. La table ronde.
(11) POHER C, (1971) - Etude statistique des rapports d'observations du phénomène OVNI.
(12) DUVAL J, (1979) - Règles de codage des témoignages d'OVNI - 4ème version GEPAN.
(13) HYNEK JA, (1972) - The UFO experience, a scientific inquiry (REGNERY) .
(14) SAUNDERS D, (1976) - Extrinsic factors in UFO reporting (com. personnelle)
(15) Mc CAMPBELL JM, t 1973) - UFOLOGY - Jaymac Company .
(16) DELECOLLE R, PAGES J, ROSPARS JP, (1978) - Codage informatique de documents, première version du livret (20 dec. 78) - Document interne GEPAN.
(17) MICHEL A, (1958) - Mystérieux objets célestes - Arthaud (réédition : Seghers 1978) .
(18) C'est à partir de ce fichier que nous avons réalisé cette étude "Analyse statistique sur les rapports d'observation du type D : rencontres rapprochées, à propos de la cohérence interne entre conditions d'observation et détails décrits : structuraux, sonores, lumineux" - GEPAN - dec. 77 –tome 1 Annexe 6.

CHAPITRE 6

LES MÉTHODES DE TRAITEMENT DE L'INFORMATION APPLIQUÉES AUX PHENOMENES OVNI

Par Michel MARCUS (Stagiaire Ecole Centrale)

1. INFORMATION

1.1. PRINCIPE

Nous avons examiné au cours du chapitre 4 la manière dont l'expert était appelé à qualifier le document ou témoignage qui lui était soumis.

Nous avons cherché dans la direction suivante "Quelle est la quantité d'information apportée par un rapport ?".

Nous avons donc cherché dans la théorie de l'information ce qui pouvait être utilisable pour notre problème. Avant de conclure sur l'intérêt de cette théorie pour le problème OVNI, notons simplement que le nom "théorie de l'information" est un peu mal choisi car c'est plutôt une théorie de la communication qu'a élaborée : SHANNON .

1.2. THEORIE DE L'INFORMATION

Définition :
Soit un évènement E qui est réalisé avec une probabilité p(E). Si on apprend que E est réalisé, on dira que l'information reçue est I = -log p(E). Choix de l'unité I = -$\log_2$ p(E) est exprimer en bit.

Modèle de la source sans mémoire Soit une source S émettant des symboles venant d 'un alphabet A = s_1, s_2, ..., s_q avec des probabilités d'occurence P (s_1), P (s_2), . . . P (s_q).

On peut calculer l'information moyenne procurée pair une source sans mémoire de la façon suivante :
Si un symbole s_i apparaît, on obtient une information I (s_i) = -log P (s_i) la probabilité que cela arrive est P(s_i) donc la quantité moyenne d'information par symbole de la source est :

$$H(s) = -\sum_{s}^{i} p(s_i) \log p(s_i)$$

sommation sur les q symboles de l'alphabet A de la source S.
On appelle cette quantité l'entropie de la source.
Propriété :

$$H(s) \leq \log q$$

Définition :

$$R(s) = 1 - \frac{H(s)}{\log_q}$$

redondance

Extension :
Soit deux sources S1 et S2 indépendantes, émettant ensemble et constituant ainsi S alors :
H(S) = H(S_1) + H(S_2)

Modèle de source de MARKOV :
Ici l'occurrence d'un symbole s_i dépend des m symboles précédents source d'ordre m :

$$H(s) = -\sum_{S^{m+1}} P(s_{j1}, s_{j2}, \ldots s_{jm}, s_i) \log(P(s_i / s_{ji}, s_{jm}))$$

Estimation de l'entropie :
On veut estimer l'entropie d'une source à partir de l'observation de N sorties.
On a :

$$\hat{H}(s) = \sum_{S} \hat{P}_i \log \frac{1}{\hat{P}_i}$$

Et :

$$E\left[H(s)\right] = H(s) - \frac{q-1}{2N} \log_2(e + O(\frac{1}{N^2}))\left[e = 2,718\ldots\right]$$

1.3. APPLICATION

Considérons le fichier informatique :

Considérons une colonne (la forme par exemple). La suite des symboles du codage de cette colonise peut être considérée comme une source d'information émettant cette suite de symboles.

En évaluant les proportions de chaque signe, on peut estimer les probabilités d'occurence et partant, la quantité d'information contenue dans un signe et la quantité d'information moyenne de la source ou en entropie.

1.4. COMMENTAIRES

Le défaut majeur de cette opération est que pour chaque colonne du fichier, un des symboles du codage note le manque d'indication de la variable. Et ce symbole est considéré comme ayant même valeur que les autres.

On pourra donc arriver au paradoxe suivant : supposons que, à un moment donné, tous les témoignages ont donné une indication du département, et qu'un témoignage arrive sans indication du département, on trouvera que ce document apporte beaucoup d'informations.

2. ÉTUDES STATISTIQUES

L'objectif est d'essayer de dégager des propriétés du phénomène OVNI, à partir de ces observations prises collectivement.
Il est entendu que ces résultats dépendront du codage.

2.1. PHOTOGRAPHIE DU FICHIER

Pour chaque critère du codage, on établit les pourcentages respectifs des modalités de ce critère.
Exemple :

Nombre de témoins : 3,8 % information non disponible
(sur 106 cas D) 2,8 % de 0 à 13 ans
13,2 % de 14 à 20 ans
68, 9 % de 21 à 59 ans
11,3 % de 60 ans et plus

Le chapitre précédent sur la théorie de l' information donne, par exemple, un autre moyen de photographier le fichier, en permettant le calcul de l'entropie de la source à partir d'une estimation basée sur les fréquences observées.
Exemple :
- Nombre de témoins :

$$\hat{H}(S) = \sum_i \hat{P}_i \log_2 \frac{1}{P_i}$$ en bits = 1,43 bits

Une autre visualisation du fichier est de considérer non plus les pourcentages mais les quantités réelles. Cela donne une suite de valeurs dont on peut extraire la moyenne, le moment d'ordre 2, d'ordre 3, etc…
Exemple :
Nombre de témoins (sur 106 cas) : x_i 4/1 . 3/2 . 14/3 . 73/4 . 12/5

$$\overline{x} = \frac{1}{n}\sum_i xi \qquad m_2 = \frac{1}{n}\sum_i \left(x_i^{\,2}\right) - x^{-2} \qquad m_3 = \frac{1}{n}\sum_i \left(x_i^{\,3}\right) - \frac{3}{n}\overline{x}\sum\left(x_i^{\,2}\right) + 2x^{-3}$$

x=21,20 m_2=689,36 m_3=25344,58

Ces diverses manipulations sont à faire à la fois pour tout le fichier des observations non-identifiées par le témoin, ainsi que pour les classes A, B, C et D, ce qui permet de caractériser le phénomène OVNI.

Tous ces nombres (%, entropie, moyenne, dispersion) sont un résumé de l'ensemble du fichier. Leur évolution dans le temps montrera l'évolution du fichier et donc des observation.

2.1.1. Age des témoins

âge :	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
obs.:	2	0	1	0	1	3	3	2	4	2	3	10	6	3	4	4	5	5	3

âge :	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47
obs.:	3	2	5	3	2	3	3	1	2	3	0	3	3	3	0	1	1	1	3

âge :	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
obs.:	4	0	2	1	0	1	4	1	4	2	1	1	0	1	0	1	2	1	2

âge :	67	68	69	70	71	84
obs.:	0	0	0	0	3	1

135 observations : $\sum x^2 = 192747$

$\sum x = 4637$

$\bar{x} = 34,34$

$s_{\text{échantillons}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = 15,80$

$s'_{\text{population}} = s\sqrt{\frac{n-1}{n}} = 15,74$

La 3ème version du codage ne donnait que les 4 modalités suivantes pour l'âge, soit :

E : de 0 à 13 ans
J : de 14 à 20 ans
A : de 21 à 59 ans
V : de 60 ans et +

La 4ème version (1) est plus précise car elle donne l'âge en clair (ou ** si inconnu).

(1) cette 4ème version est celle explicitée au chapitre 2.

2.1.2. Heure de l'observation

Le codage dans le fichier est le suivant :

heure	h h m n e |estimation
locale	0 2 3 0 Z |

CODE :
16 - M : matin
4 - P : après midi
10 - C : crépuscule
50 - Z : vers minuit
3 - L : aurore
3 - V : Vers midi
4 - S : Soirée
15 - D : début de la suit
18 - F : fin de la nuit

COMMENTAIRES :
Les règles ne donnent pas le passage heure → estimée. En regardant le fichier, on trouve par exemple :
2230D et 2230Z.
On peut penser que la date y est pour quelque chose, effectivement le 1er se trouve en juin, l'autre en octobre. Dans la mesure où l'heure est indiquée tout le temps, on doit trouver une certaine corrélation entre mois, heure et estimée, ou mis autrement, il y a redondance entre ces données. Cette redondance est inutile et se reportant à un travail sur les diagnostics automatiques, cela serait même préjudiciable à une classification automatique.

Pour la représentation graphique, à part l'histogramme que nous avons présenté, on peut penser à un diagramme "polaire" (cf. sémiologie graphique BERTIN).

3. ANALYSE DE DONNEES

3.1. ANALYSE MULTIDIMENSIONNELLE

Citons les résultats d'une étude conduite par le GEPAN. Les données étaient constituées par 736 observations d'OVNI, qui ont été mises sous forme d 'un tableau homogène de 736 lignes et 219 colonnes.

* Premier traitement :
Construction d'un tableau croisant la description des conditions de l'observation (111 modalités) et la description de l'observation (108 modalités) .

* Résultats :
- le nuage est assez sphérique ;
- l'angle entre les hypers plans : condition et description, est proche de 90° ;
- le 1er axe oppose les évènements bien décrits aux évènements peu décrits ;
- le 2ème axe semble lié aux conditions d'observation.

* Conclusions :
Les tendances observées apportent peu de révélations et permettent difficilement de dresser une typologie.

* Deuxième traitement :
Construction d'un tableau (124 x 124) croisant les descriptions des observations avec elles-mêmes.

* Résultats :
- le nuage est assez sphérique ;
- l'angle des deux hypers-plans est proche de 70° (liaison globale plus forte que précédemment) ;
- le 1er axe factoriel s'interprète facilement comme étant la distance d'observation ;
- le 2ème axe s'interprète comme étant la durée de l'observation.

3.2. CONCLUSION

Il semble que la méthode d'analyse utilisée permette de mettre en évidence certains éléments structuraux.

Il serait donc approprié de l'utiliser aussi pour les objets identifiés.

En premier lieu, son élaboration et amélioration tiendraient compte des résultats des expertises des objets ce qui garantirait en second lieu son emploi pour les objets non-identifiés, et enfin donnerait peut-être un outil pour une expertise plus ou moins automatique, chaque observation arrivant étant prise comme un élément supplémentaire.

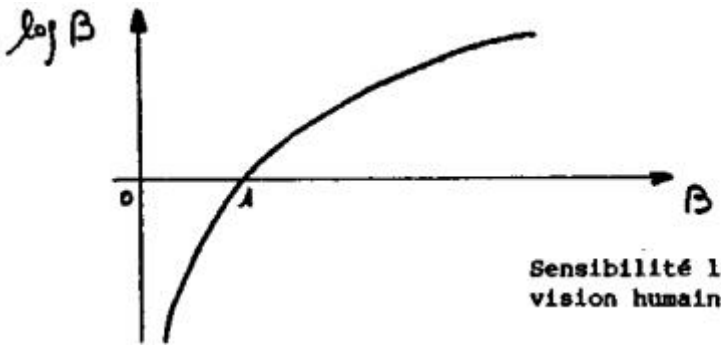
4. LES ESSAIS DE MODÉLISATION

4.1. PERCEPTION

Nous allons présenter ici une étude de C. POHER dont le but était de montrer que les observations d'OVNI ne correspondaient pas a des hallucination ou canulars. Nous avons changé quelques détails mineurs dans cette étude.

Plus un objet est lumineux, plus il est probable qu'il soit perçu. Si toutes les luminosités sont représentées uniformément, une observation prise au hasard est plus probablement celle d'un objet très lumineux que peu, ou, des classes d'objets lumineux rangées par proportion croissants donnent une échelle de luminosité croissants.

Il faut donc trouver un moyen de passer d'une échelle de proportion à une échelle de luminosité (ou brillance).



Sensibilité logarithmique de la vision humaine

Pour une classe d'objets représentant toutes les luminosités : B varie de 0 à + l'infini
Log B varie de - infini à + infini -

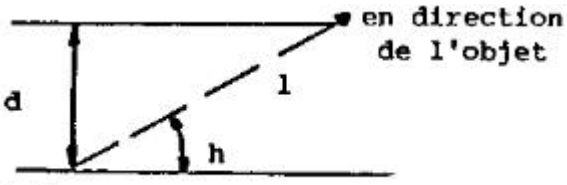
Pour une classe d'objets représentant uniquement des luminosités faibles :
B varie de 0 à 1 (par exemple)
Log B varie de – l'infini à 0

On vent pour ces deux classes, pouvoir passer des proportions de sous-classes constituant la classe entière à des valeurs sur l'axe "log B".

- 1ère classe :
  On veut passer de 0 ≤ proportion ≤ 1 à - infini ≤ log B ≤ + infini
  Les fonctions y = x log (x/(1-x)) effectuent ce passage.

- 2ème classe :
  On veut passer de 0 ≤ proportion ≤ 1 à - infini ≤ log B ≤ 0
  Les fonctions y = λ log x effectuent ce passage.

OBTENTION DE LA LOI DE BOUGUER :



Témoin

épaisseur de l'atmosphère : d l=d/sin h

Décroissance exponentielle de l'intensité lumineuse à travers un milieu d'épaisseur x : $B = B_o e^{-kx}$

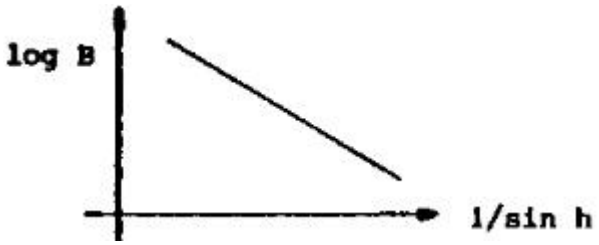
Donc ici : $B = B_o e^{-kl} = B_o e^{-kd/\sin h}$

log B = - (kd/sin h) + log B_o k et d étant sensiblement constant

$\log B = -(C_1/\sin h) + C_2$

Cette loi donne donc la variation de la magnitude en fonction de la hauteur angulaire d'un même objet vu d'endroits différents.

LOI DE BOUGUER :
Condition d'application : astronomie - étoiles
Expression :



Comment vérifier cette loi ? :

- Choisir parmi les observations celles dont les caractéristiques se rapprochent au mieux des conditions d'applicabilité de la loi ;

- Partitionner les observations en classes de hauteur angulaire, calculer les portions de chaque classe, effectuer le passage de ces proportions à une échelle de luminosité, et porter les résultats dans le plan "plan de BOUGUER" ainsi redéfini.

Réalisation du test :

- Choix des valeurs de h dans chaque classe de hauteur angulaire,
- Choix des critères à retenir dans le fichier d'observation.

Hauteur angulaire moyenne :

*	pas d'indication
1	de 0 à 15°
2	de 15 à 30°
3	de 30 à 45°
4	de 45 à 60°
5	de 60 à 90°
6	au-dessus de l'horizon (avion)
7	avion ou au-dessus
8	au sol ou près du sol
.	autres types

Dans le cas présent, on s'intéresse aux modalités :
1 U 2 U 3 U 4 U 5. On a donc deux modalités d'emploi :
- moyenne,
- borne supérieure.

Moyenne :

CLASSES	MOYENNE	1/sin h
1	7° 30'	7,66
2	22° 30'	2,61
3	37° 30'	1,6
4	52° 30'	1,26
5	75°	1,03

Borne supérieure :

CLASSES	MOYENNE	1/sin h
1	15°	3.86
2	30°	2
3	45°	1,41
4	60°	1,15
5	90°	1

Critères :

- CR 5 : type de témoignage : 1 lumière nocturne
- CR 17: distance minimale : 5 et 6 - **≥** 1 km
- CR 18 : méthode observation :
  - A : oeil nu au sol
  - B, C et D : instruments optiques
- CR 20 : forme objet principal : K - ponctuelle
- CR 23 : couleur : toutes 1 à 1
- CR 28 : hauteur angulaire : 1, 2, 3, 4, 5 - de 0° à 90°

Autres actions :

- appliquer le test aux cas (A U B) ; si échec, chercher pourquoi et redéfinir le test ;

- la loi de BOUGUER lie deux grandeurs, sa vérification peut permettre de ranger plus quantitativement les modalités du critère 22, luminosité à partir d'un test basé sur des probabilités d'observation liées aux volumes délimités par les classes de hauteur angulaire, sachant que la vérification du test est assurée.

COMMENTAIRE

1) il y a une erreur dans le raisonnement de POHER, c'est d'appliquer une loi valable pour les objets situés à l'infini. Dans la mesure où les observations sort faites dans l'atmosphère, le seul facteur d'extinction est la distance entre le phénomène et le témoin.

   Avec : d = h/sin θ

2) Il semble plus intéressant et susceptible d'ouvrir des voies d'études plus fructueuses, de considérer ce travail comme un essai de modélisation du phénomène en tenant compte d'un modèle de la perception humaine, que comme une preuve de l'existence objective des OVNI.

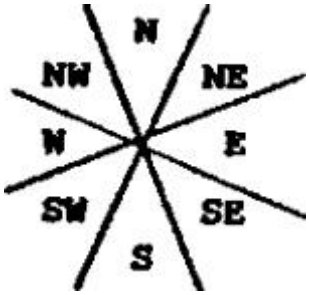
4.2. MODELISATION SPATIALE

4.2.1. Hypothèse

Le phénomène se produit de façon équiprobable dans l'espace, c'est ce que nous allons essayer de vérifier, suivant plusieurs modalités.

4.2.2. Direction azimutale

On peut penser que les volumes engendrés par les directions azimutales sont égaux, on devrait obtenir des nombres égaux d'observations:



4.2.3. Hauteur angulaire

Les hauteurs angulaires sont codées ainsi :

- de 0° à 15°
- de 15° à 30°
- de 30° à 45°
- de 45° à 60°
- de 60° à 90°

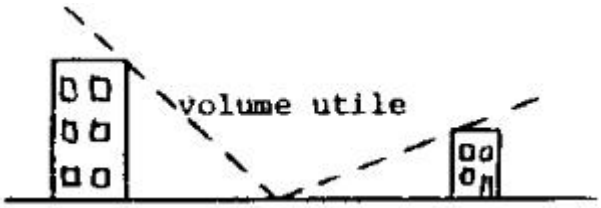
Ces classes délimitent des volumes. Ainsi, pour la distance, il faut décider comment choisir les limites du volume.

Exemple :

- ciel clair : a priori sphérique - secteur sphérique angulaire
- ciel nuageux : ici problème de définition du volume pour les hauteurs entre 0 et 15°

Autre exemple :

- campagne : quelques obstacles : arbres
- ville : maisons, immeubles. ici, un modèle des obstacles en milieu urbain et rural serait utile (cf. les objets fractals - MANDELBROT).



1er cas :

Volume secteur angulaire sphérique : $V_{\alpha\beta} = \frac{2\pi R^3}{3}\left(\sin\beta - \sin\alpha\right)$

0. 15 : 0,26
15. 30 : 0,24
30. 45 : 0,21
45. 60 : 0,16
60. 90 : 0,13

2ème cas :

Volume secteur angulaire conique : $V_{\alpha\beta} = \frac{1}{3}\pi H^3\left(\cot g^2\alpha - \cot g^2\beta\right)$

0. 15 : infini
15. 30 : 10; 93
30. 45 : 2
45. 60 : 0,66
60. 90 : 0,33

Prenons par exemple : H = 10 000 m

D = H/sin15° # 40 km

RESULTATS :

Hauteur angulaire	Résultats	Fréquences prévues	Nombres prévus
0 - 15	24	0,26	17,16
15 - 30	11	0,24	15,84
30 - 45	12	0,21	13,66
45 - 60	7	0,16	10,56
60 - 90	12	0,13	8,58

Xhi^2 =7.02 à 4° de liberté

Direction azimutale	Résultats	Fréquences prévues	Nombres prévus
Nord	10	0,125	11,88
Nord-est	12	0,125	11,88
Est	14	0,125	11,88
Sud-est	17	0,125	11,88
Sud	10	0,125	11,88
Sud- ouest	9	0,125	11,88
Ouest	7	0,125	11,88
Nord-ouest	16	0,125	11,88

Xhi^2=7,31 à 7° de liberté.

4.2.4. Distance

La distance minimale d'observation est codée de la façon suivante :

- 3 cases pour la distance chiffrée
- 4ème case pour la classe :
  - M : mètres (de 0 à 999 m)
  - R : kilomètres (de 1 à 3 km)
  - A : $\geq$ 3 km

Exemple donné :
- 55 mètres 055M
- 1,5 km 1.5K

Dans le fichier, on trouve :
- 005K ..... contrairement au codage
- 001K et 999M ..... quelle est la différence ?
- 010A ..... conformément au codage

Il s'agit de retrouver ensuite la mesure de la distance (en mètres par exemple) et d'affecter cette observation à une classe de distance pour effectuer un test en Xhi^2.

Il est évident à ce stade, qu'il s'agirait de faire intervenir la perception humaine pour tenir compte des limitations de celle-ci. C'est pourquoi nous allons présenter quelques notions de psychologie tirées de réf . PIERON.

4.3. PSYCHOLOGIE OBJECTIVE

Seuil

Point où la valeur croissante d'un stimulus commence à exercer un effet sensoriel déterminé :

- éveiller une sensation (seuil absolu)
- entraîner une modification discriminante d'une sensation (seuil différentiel successif).

Conception de FECHNER

FECHNER, à partir des résultats expérimentaux indiquant la constance de la valeur relative des accroissements juste perceptibles avait conclu à la validité d'une loi obtenue en faisant un passage à l'infiniment petit de :

Δ perçu = k (Δ à percevoir)/ (à percevoir)

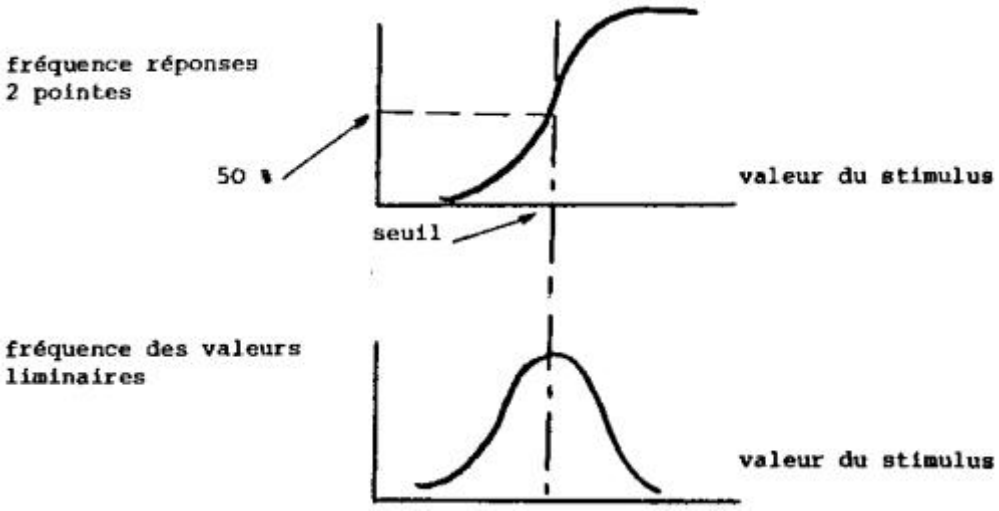
puis en intégrant l'équation différentielle obtenue, ce qui donne :

perçu = k log (à percevoir) + h

Mesure du seuil par la méthode constante

Le seuil est défini comme étant la valeur qui a autant de chances de ne pas être atteinte que d'être dépassée. Mesure du seuil discriminatif spatiale d'acuité tactile : 8 écarts équidistants ont été utilisés, chacun 20 fois.

On obtient :



Par exemple, dans l'étude POHER, il convient de choisir des objets lumineux ponctuels, de mouvement & de forme anodins. D'un point de vue psychologie de la perception, faire des hypothèses comme la précédente c'est essayer de répondre à la question : "qu'est-ce qui fait qu'on remarque quelque chose ?" (ici dans le ciel).

Il faudra alors faire une hypothèse sur les grandeurs proposées à la perception du témoin par exemple pour l'étude de POHER : une luminosité unique, des luminosités équiprobables, etc...

Et, dans la mesure où l'on fait l'hypothèse qu'une grandeur physique est responsable de la probabilité de vision d'un objet, Il faut faire en sorte de choisir pour tester cette hypothèse, les observations où seule intervient cette grandeur.

4.4. LOI DE HARTMANN

Dans le rapport CONDON, une étude a été publiée par le Dr. HARTMANN sur le processus de perception, d'interprétation et d'émission de rapports d'observations.

Le modèle élaboré par HARTMANN conduit à une loi (dite : loi de HARTMANN) :

$$\mathbf{N_r = C_r \, . \, N \, . \, t}$$

N_r = Nombre d 'obs.
N = Nombre d'habitants survolés
t = durée du survol
C_r = constante évaluée par HARTMANN entre $1/3{,}5.10^{7}$ et $1/1{,}5.10^{6}$ homme/s

POHER dans une étude a estimé cette constante pour la FRANCE, comme étant de l'ordre de 1 à 2.10^{-6}. Nous avons voulu vérifier cette loi sur le fichier Gendarmerie du GEPAN (cas D) en utilisant une répartition par département.

La répartition est indiquée ci-dessous elle ne permet pas de vérifications de l'ordre de celle qu'on voulait faire. L'intérêt de travailler sur le fichier GEPAN est qu'il a une certaine homogénéité mais il n'est pas actuellement assez étendu pour certaines études.

Rapport du nombre d'observations en fonction des départements français :

Nbre d'obs.	Numéros des départements
0	1 . 3 . 5 . 12 . 14 . 18 . 20 . 22 . 23 . 24 . 25 . 27 . 36 . 37 . 40 . 43 . 44 . 46 . 48 . 56 . 57 . 63 . 65 . 70. 74 . 75. 84 . 90 . 91 . 92 . 93 . 94 . 95 . 78
1	2 . 4 . 6 . 9 . 10 . 15 . 16 . 19 . 30 . 31 . 32 . 35 . 39 . 42 . 47. 49 . 50 . 53 . 58 . 61 . 64 . 66 . 67 . 72 . 77 . 79 . 80 . 81 . 86 . 87 . 89
2	7 . 8 . 11 . 13 . 26 . 28 . 33. 45. 51. 60 . 68. 73 . 85
3	34 . 52 . 69 . 82 . 83 . 88
4	17. 29 . 34 . 62
5	38 . 55
6	
7	59

5. COMMENT ÉTUDIER UNE HYPOTHESE : PLAUSIBILITE ET CONFIRMATION

5.1. PLAUSIBILITE

Nous citerons ici un article de POLYA, auquel nous avons déjà fait référence dans le rapport de séminaire : "Intelligence artificielle et recherche heuristique (MA 78.79) ". (Ecole Centrale).

5.1.1. probabilité

Point de vue objectif : ex. 1300 garçons/2500 naissances --> fréquence 0,52. On cherche à prédire les fréquences observables à partir des fréquences observées.

Point de vue subjectif : mesurer la croyance en telle ou telle chose. Par exemple : véracité d'un théorème.

5.1.2. Degré de proximité - degré de croyance

Chaque étape de la recherche s'accompagne d'un changement d'attitude du chercheur.

Proximité : intensité éveillée par un problème donné, plus espoir d'en trouver la solution.

Croyance : commence-t-on à chercher si T est vrai ou faux ?

5.1.3. Application des probabilités

La probabilité est mesurée par un nombre déterminé entre 0 et 1. A une plausibilité, on fera correspondre une variable entre 0 et 1 (ou nombre indéterminé).

Véracité d'un théorème :

P_o (T) : plausibilité puis P_1(T) , P_2 (T) , . . . etc. . .

Si aucun effet :

$P_o(T) = P_1(T) = P_2(T)$

si T réfuté : P (T) = 0

si T démontré : P (T) = 1

5.1.4. Exemples d'application

- On cherche à démontrer T
  (H ⇒ T) vrai
  on trouve H faux

  $$p(T) - p_1(T) = p(H) \cdot p(\bar{T}/\bar{H})$$

  Conclusion : $p(T) > p_1(T)$
  l'écart $p(T) - p_1(T)$ est d'autant plus grand que $p(H)$ ou $p(\bar{T}/\bar{H})$ est grand

- On cherche à démontrer T
  (T ⇒ C) vrai
  On trouve C vrai

  $$\frac{p_1(T)}{p(T)} = \frac{1}{p(C)}$$

  Conclusion : $p_1(T) > p(T)$

  $\frac{p_1(T)}{p(T)}$ est d'autant plus grand que p(C) est petit

- On cherche à démontrer T
  $T \Rightarrow C_1$ et C_2 et .... C_n...
  On trouve successivement C_1, C_2, C_n

  $$\frac{P_n(T)}{P_{n-1}(T)} = \frac{1}{P_{(C_n/C_1\ C_2 \ldots C_{n-1})}}$$

Conclusion :

- 1er cas : C_n : conséquence logique des n-1 premières vérifications
  $\mathbf{P}(C_n/C_1\ C_2.....C_{n-1}) = 1$ et $P_n(T) = P_{n-1}(T)$

- 2ème cas : C_n : nouvelle conséquence $P_n(T)$ $P_{n-1}(T)$ et moins C_n est une conséquence des n-1 C_i, plus $\mathbf{P}(C_n/C_1\ C_2.....C_{n-1}) \approx 0$ et plus $P_n(T) \geq P_{n-1}(T)$

5.1.5. Commentaire

Il manque les moyens d'évaluer la plausibilité d'un phénomène, les lignes précédentes montrant seulement l'évolution de cette plausibilité.

5.2. CONFIRMATION

Nous citerons ici un article de W.C. SALMON "confirmation" Scientific American - Mai 73 :

Pendant des siècles, des chercheurs ont testé, confirmé et réfuté des hypothèses par l'observation et l'expérience. Pourtant, la logique du processus est encore loin d'être comprise.

A - Paradoxe du corbeau noir (Carl G. HEMPEL)

Des observations de corbeaux noirs (en l'absence d'observation de corbeaux d'autres couleurs) seront généralement prises comme des confirmations de la généralisation "tous les corbeaux sont noirs". Cette assertion est logiquement équivalente à "toutes choses non

noires sont des non-corbeaux". Donc l'observation de choses noires par exemple "un vase vert" confirmera cette deuxième assertion et par suite la première.

B - "les porcs ont des ailes" (RUSSEL)

Considérons l'hypothèse "les porcs ont des ailes". En conjonction avec le fait observé (condition initiale) que le porc est bon à manger, nous déduisons la conséquence - nous prédisons - que des créatures ailées sont bonnes à manger. Quand nous voyons que des personnes apprécient le canard, ou la dinde, nous observons que la conséquence - ou prédiction - est vraie ; il semble donc que nous ayons eu une confirmation de l'hypothèse de départ.

C - "Blert-veu" paradoxe

Deux couleurs particulières sont définies. Considérons un point t_o dans le futur (31.12.2000). Un objet sera veu pendant la période s'étendant sur le 20ème et 21ème siècle, s'il est vert pendant le 20ème et change au bleu au 21ème. Blert est défini de façon analogue : un objet sera blert pendant la période s'il est bleu avec t_o et vert après. Normalement nous dirions que l'observation d'émeraudes vertes (en l'absence d'observation, d'autres émeraudes d'autres couleurs) tend à confirmer la généralisation "toutes les émeraudes sont vertes". Puisque la date actuelle est avant la date (31.12.2000) l'émeraude observée comme étant verte est aussi observée comme étant veu et confirme ainsi l'hypothèse "toutes les émeraudes sont veu". Comment alors devrions-nous prédire quant aux émeraudes du 21ème siècle ?

D - Scientifiques et micro-organismes

Puisqu'il est très probable que prenant un scientifique parmi tous ceux qui ont vécu et vivent, que ce soit un contemporain (on estime que 90 % de tous les scientifiques sont vivants aujourd'hui) et puisque il est très probable que prenant un organisme vivant aujourd'hui, ce soit un microorganisme, alors étant donné que SMITH est un scientifique, il est probable qu'il soit un micro-organisme.

PROBABILITE

Le meilleur antidote contre les erreurs qui viennent de l'intuition en étudiant la confirmation est de se rapporter à la théorie des probabilités. Si la confirmation peut être vue comme un type de probabilité, il est facile de montrer que les relations de confirmation ne sont pas transitives. Il peut aussi être montré que la vérité d'une conséquence d'une hypothèse n'augmente pas nécessairement la probabilité de cette hypothèse (POLYA parle de plausibilité). Il faut aussi lever une ambiguité qui existe entre la confirmation "incrémentielle" et la confirmation "absolue". En effet, on dira que la théorie de la relativité restreinte a été confirmée "incrémentièllement" par expérience (les horloges) et est confirmée "absolument" par le corps des évidences la supportant. La confirmation "incrémentielle" par exemple, a des propriétés inhabituelles.

NON CONFIRMATION

Il faut aussi étudier le cas où le résultat d'une expérience est négatif. Ceci ne peut pas être pris comme la réfutation de l'hypothèse puisque dans l'expérience sont ajoutées d'autres hypothèses pour la réaliser. Considérons l'exemple historique suivant : "Les fausses prédictions de la théorie de NEWTON concernant la trajectoire d'URANUS n'ont pas réfuté la physique de NEWTON mais ont mené à la découverte de SATURNE. Par contre, les irrégularités de l'orbite de MERCURE par rapport à la physique Newtonienne n'ont pas

mené à la découverte d'une nouvelle planète, mais plutôt à une mise en question de la mécanique newtonienne puisque elles furent une preuve pour la mécanique d'EINSTEIN."

MOYENS

La position du chercheur actuellement ressemble à celle des mathématiciens qui n'avaient pas formalisé la logique qu'ils employaient. Les paradoxes de RUSSEL ont provoqué une crise dans les fondements des mathématiques. Deux voies semblent permettre des résultats dans l'étude de paradoxes de confirmation. D'abord le théorème de BAYES en probabilité, et enfin la reconnaissance du concept de confirmation incrémentielle pour définir une mesure au moyen de laquelle la confirmation incrémentielle sera étudiée.

"tous" et "presque tous" peuvent être très différents. Si tous les A sont des B et tous les B des C, il est vrai que tous les A sont des C (gauche). Pourtant bien que presque tous les scientifiques A soient vivants (8) et que presque tous les vivants (B) soient des micro-organismes (C), aucun scientifique n'est un micro-organisme.

6. CONCLUSION

Le GEPAN a choisi d'étudier le phénomène OVNI d'une façon scientifique. Pour citer JP. ROSPARS, disons qu'il n'y a pas de sujet indigne de la science, mais que des méthodes indignes d'elle. Le problème du GEPAN est donc un problème de méthodes.

Le point crucial dans les actions du GEPAN est évidemment l'expertise. La mise au point "d'instruments de mesure" semble nécessaire pour relayer l'expert dans son travail. Cela est particulièrement frappant dans le cas de contactés avec messages.

La menée de l'étude statistique, de modélisation, ou de description de données sera facilitée par des outils informatiques faciles à utiliser, ne nécessitent que la définition de quelques paramètres pour que le traitement soit réalisé.

Il est entendu que ces traitements seront d'autant plus validés, appliqués à l'ensemble ou à des sous-ensembles des objets identifiés, ils "collent avec l'expertise.

Il sera intéressant de suivre les développements d'études de sujets extérieurs au phénomène OVNI, intéressants soit parce que leurs résultats pourront être inclus dans les études du GEPAN, soit parce que des outils nouveaux seront développés et auront leur place au GEPAN.

Dans le même ordre d'idée, des tentatives de formalisation des étapes de la méthode scientifique, seraient utiles pour le GEPAN dans d'éventuels discours épistémologiques.

CONCLUSION

par Alain ESTERLE

Nous avons vu dans les chapitres 1, 2 et 3 comment s'articulait la démarche du GEPAN pour saisir les données, les expertiser, les classer, les coder et en tirer une image statistique.

Les techniques mises en oeuvre ne doivent en aucun car être considérées comme une méthodologie définie une fois pour toutes conduisant nécessairement à une conclusion correcte, et ceci pour deux raisons :

- dune part, des difficultés pratiques et logiques apparues dans l'utilisation de cette démarche,

- d'autre part, une réflexion générale a permis de mieux comprendre les fondements de ces difficultés et la possibilité de développer d'autres approches (chapitres 4, 5 et 6).

La démarche utilisée a donc permis de catalyser la réflexion sur la méthode autant que sur son objet. Mais, au stade préliminaire où nous en sommes, il n'était pas possible d'en attendre plus.

Ceci signifie que nous sommes à la veille d'une redéfinition globale des principes du pré-traitement non plus inspirée seulement des techniques antérieures (classification, expertise, codage) mais trouvant sa source dans la spécificité même du problème traité et l'originalité des données disponibles.

Cette redéfinition est actuellement en cours. Après une phase d'expérimentation et d'ajustements, qui prendra quelques mois, cette nouvelle méthode de pré-traitement sera décrite en détail dans une prochaine note technique.