

CENTRE NATIONAL D'ETUDES SPATIALES

TOULOUSE, -- LE 25 MARS 1981 --
N° 085 CT/GEPAN

NOTE TECHNIQUE

N°4

Recherche statistique d'une typologie des descriptions de Phénomènes Aérospatiaux Non identifiés

GROUPE D'ÉTUDES
DES PHÉNOMÈNES AÉROSPATIAUX NON IDENTIFIÉS



SOMMAIRE

Introduction

1. MISE EN OEUVRE DES ANALYSES

1.1. LES OUTILS

- 1.1.1. Modèle
- 1.1.2. Analyses factorielles
- 1.1.3. Classification automatique

1.2. PREPARATION DES DONNEES

- 1.2.1. Localisation
- 1.2.2. Conditions sociologiques
- 1.2.3. Conditions physiques
- 1.2.4. Description

2. DESCRIPTION DES VARIABLES ET DE LEURS MODALITÉS

2.1. DESCRIPTION DES VARIABLES

2.2. DESCRIPTION DES MODALITES

- 2.2.1. Localisation
- 2.2.2. Conditions sociologiques
- 2.2.3. Conditions physiques
- 2.2.4. Estimation de la distance
- 2.2.5. Description de l'observation
 - . Préliminaire
 - . Comportement du phénomène
 - . Description du phénomène
 - . Description conjointe du phénomène et de son comportement
- 2.2.6. Liaisons inter-groupes

2.3. RESUME DE L'ANALYSE DESCRIPTIVE DES VARIABLES ET DE LEURS MODALITES

3. CLASSIFICATION AUTOMATIQUE DES OBSERVATIONS

3.1. PRÉLIMINAIRE

3.2. RESULTATS

- 3.2.1. Comportement du phénomène
- 3.2.2. Description du phénomène
- 3.2.3. Description des phénomènes et de leurs comportements

4. CONCLUSION

4.1. LES TECHNIQUES D'ANALYSES

4.2. LE CODAGE

4.3. CONCLUSION DE CETTE ETUDE

4.4. ... A SUIVRE....

BIBLIOGRAPHIE

INTRODUCTION

Le travail poursuit la démarche abordée en (BES-80) à l'aide d'outils statistiques élémentaires (histogrammes - tests d'ajustement) en utilisant cette fois les méthodes classiques mais plus élaborées, de l'analyse des données multidimensionnelles. Rappelons que ces outils permettent une approche synthétique d'un fichier donné (il s'agit ici des témoignages recueillis par la Gendarmerie nationale), mais n'autorisent aucune inférence sur l'ensemble des phénomènes.

Alors que l'étude précédente n'envisageait que la description de chacune des variables, considérées indépendamment, les outils statistiques, brièvement décrits dans le premier paragraphe, visent cette fois à mettre en évidence les liaisons (ou "proximités" ou "corrélations") présentes dans les données du fichier relatives aux observations de phénomènes aérospatiaux non identifiés (cas classés D).

Dans cette perspective, trois ensembles sont successivement étudiés:

- les variables,
- les modalités de celles-ci,
- les observations.

Comme précédemment, ce travail sans résoudre des questions fondamentales, définit ce que seront les prochaines étapes de l'étude en tâchant de cerner au mieux les limites des outils statistiques.

1. MISE EN OEUVRE DES ANALYSES

1.1. LES OUTILS

Les fondements théoriques des méthodes employées sont abondamment décrits dans la littérature (CAP-76 ; SAP-78 ; DAP-76 ; LMT-77 ; ...) avec des optiques (géométriques, matricielles, probabilistes, ...) qui diffèrent selon les habitudes des auteurs. Seuls les principes généraux de ces techniques sont brièvement rappelés ; les logiciels sont ceux diffusés par le CESIA (CEntre de Statistiques et d'Informatique Appliquées) et décrits en (LMT-77).

1.1.1. Modèle

On considère un ensemble de n individus (les rapports d'observations) décrits par p variables numériques ou non. Pour permettre les comparaisons des variables et ainsi les analyses, on commence par recoder toutes les données en des tableaux (dits disjonctifs complets) comprenant les valeurs 0 ou 1 suivant la réalisation ou non d'un événement, l'appartenance ou non à une classe.

Ainsi, par exemple, la variable "département" est remplacée par 6 variables (dites indicatrices des modalités) :

$$(1_{S-W}, 1_{S-E}, 1_{\text{centre}}, 1_W, 1_E, 1_N)$$

qui prennent des valeurs 0 ou 1 suivant la région dans laquelle s'est produite l'observation.

L'ensemble des méthodes utilisées repose alors sur une représentation vectorielle des données :

- chaque modalité d'une variable est ainsi représentée par un vecteur de R^n constitué par une suite de 0 ou de 1 suivant l'appartenance ou non de l'individu i ($i = 1, \dots, n$) à la classe en question ;
- inversement, chaque individu est également considéré comme un vecteur comportant cette fois m composantes (m est le nombre total de modalités) 0 ou 1 suivant les modalités j ($j = 1, \dots, m$) que prend l'individu en question.

1.1.2. Analyses factorielles

Les données ainsi modélisées, l'analyse vise alors à une représentation optimum, c'est-à-dire qui modifie le "moins possible" les structures de ces données. Dans ce but, les vecteurs (individus ou variables) sont projetés sur les plans engendrés par les axes principaux d'inertie du nuage** (ce sont aussi les directions associées aux plus grandes "variances"). On obtient ainsi une représentation "au mieux" des "proximités" entre individus ou entre modalités.

* L'emploi des guillemets signifie ici que l'explication précise du sens du terme se trouve dans la bibliographie citée.

** Les individus (et les variables) étant munis de poids, et les espaces de métriques, appropriés.

L'étude directe des variables (représentées par des paquets d'indicatrices) se fait en considérant non plus des vecteurs, mais les sous-espaces vectoriels (engendrés pour chacun des paquets d'indicatrices) et les "angles" que ceux-ci admettent entre eux deux à deux (les cosinus au carré de ces angles sont les coefficients T^2 de TCHUPROV - cf. CAP-76). ***

*** Par construction, ces coefficients sont en fait très faibles et leur interprétation ne fait pas l'unanimité parmi les auteurs. Il faut donc se montrer très prudent dans leur emploi.

1.1.3. Classification automatique

La méthode employée est celle des "nuées dynamiques", elle consiste à exécuter l'algorithme itératif ci-dessous afin de répartir en k classes (k fixé a priori) n points d'un espace euclidien.

- a) Sélection a priori de façon aléatoire de k groupes de quelques points comme "noyaux" provisoires des k classes.
- b) Chaque point est rangé dans la classe dont le noyau est le plus "rapproché".
- c) Les noyaux sont remplacés par d'autres points choisis autour du "centre de gravité" de chaque classe.
- d) Les points b et c sont itérés jusqu'à la stabilité des centres de gravité.

1.2. PRÉPARATION DES DONNÉES

Les outils d'analyse employés nécessitant des données de nature homogène, il faut donc procéder à un recodage à partir des descriptions sommaires obtenues précédemment (cf. BES-80) et surtout au regroupement des modalités d'effectifs trop faibles.

Ces préliminaires conduisent à l'abandon temporaire du paramètre "estimation de la taille" inexploitable dans l'immédiat car code de façon trop disparate (taille métrique, taille angulaire, taille par comparaison, ...). Cette lacune sera comblée ultérieurement dans les études plus spécifiquement orientées vers la psychologie de la perception.

A la suite de ce travail, on considérera donc 25 variables toujours réparties en quatre sous-groupes et décrivant 207 observations des années 74 à 78. Ces variables prennent les modalités présentées ci-dessous.

1.2.1. Localisation

N°	Variable	Nbre modalités	Liste des modalités
1	année	5	1974 - 75 - 76 - 77 - 78
2	mois	12	Janvier, février, ..., décembre
3	estimation heure	10	Inconnue, matin, vers midi, après-midi, soirée, crépuscule, début nuit, fin nuit, vers minuit, aurore.
4	région	7	inconnue, DOMTOM ou étranger, Sud-Ouest, Sud-Est, Centre, Ouest, Nord, Est.

1.2.2. Conditions sociologiques

N°	Variable	Nbre modalités	Liste des modalités
5	Nature du lieu	7	Inconnue, autre - désert, haute montagne, mer - habitation isolée - hameau, petit village - bourgade, banlieue - ville - très grande ville
6	Nombre de témoins	5	inconnu - un - deux - trois -- quatre et plus
7	Catégorie socio-profes.	10	inconnue - agriculteur, salarié agricole - patron industrie et commerce - prof. libérale, cadre sup. - cadre moyen - employé - personnel de service-ouvrier - autre (armée, police) - non actif -
8	Tranche d'âge	5	inconnue - enfant (< 13 ans) - adolescent (14 à 20) - adulte (21 à 60) - 60 ans et +
9	Sexe	2	masculin - féminin

1.2.3. Conditions physiques

N°	Variable	Nbre Modalités	Liste des modalités
10	Conditions météo	7	inconnue - ciel limpide -nuages épars - couvert en altitude - ciel bas sans pluie - pluie, neige - autre
11	Durée	6	inconnue - inf. à 10 s - de 10 à 59 s - de 1 à 19 mn - de 20 à 59 mn - 1 heure et +
12	Distance	5	inconnue - en mètres - en hectomètres - en kilomètres - très grande distance
13	Méthode d'obs.	7	inconnue - oeil nu au sol - avec instrument - trace physique - à bord d'un avion ou d'un bateau - voiture en mouvement - voiture à l'arrêt

1.2.4. Description

N°	Variable	Nbre Modalités	(*)	Liste des modalités
14	Nombre objets	4	obj.	un deux trois - quatre et plus
15	Forme de l'objet princ.	12	Fr.	inconnue _ disque, soucoupe - rond, circulaire sphérique - cigare, cylindre -oeuf, ovale, ovoïde - conique, toupie - carré, parallélépipède - canotier, couronne - ponctuel – croix - croissant - autre
16	Luminosité princ.	9	1p.	inconnue faible - lumineux – brillant intense – non lumineux - clignotant – non lumineux avec faisceaux - autre
17	luminosité secon.	6	1s.	sans luminosité secondaire - plus ou moins lumineux - intense – variable non périodique clignotant - autre
18	couleur	12	cl.	inconnue - rouge - orange - Jaune - vert - bleu - blanc - gris - métallique – plusieurs couleurs - couleur changeante - autre
19	caractère princ. de la trajectoire	6	tp.	inconnue, autre – droite, légère courbe - virages, arabesques – complexe mais analysable - stationnement près du sol - atterrissage
20	Caractère sec. de la trajectoire	9	ts.	sans caractère secondaire - linéaire _ virages, arabesques - complexe mais analysable - station près du sol - atterrissage - monte vers les étoiles - vitesse nulle puis lente - vitesse nulle puis rapide
21	estimation vitesse	7	v	inconnue - lent, immobile - très rapide - variable - fulgurante - comme un avion - autre
22	estimation accélération	5		inconnue - faible - variable - élevée - autre
23	bruit	5	br.	inconnu - silence - plutôt dans les graves - plutôt dans les aigus - autre
24	hauteur angulaire (début)	8		inconnue - 0 à 15° - 15 à 30° - 30 à 45 - 45 à 60° - 60 à 90 - vu d'avion - vu au sol
25	azimut (début)	10		inconnu - au zénith - nord - nord/ est – est – sud/est - sud - sud/ouest - ouest - nord/ouest

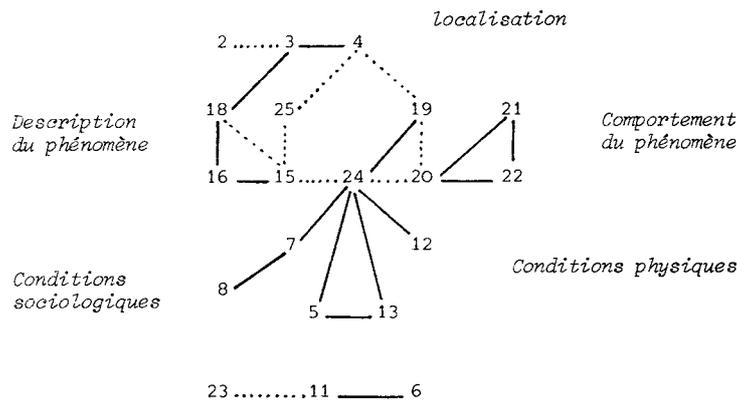
(*) Abréviation utilisée pour les figures

2. DESCRIPTION DES VARIABLES ET DE LEURS MODALITÉS

D'une manière générale, pour ce travail, on note que les liaisons rencontrées sont faibles (cf. tableau 1) et seules des tendances relativement peu marquées peuvent être relevées à partir des données étudiées.

2.1. DESCRIPTION DES VARIABLES

Une première approche élémentaire du corpus des données consiste à construire à partir du tableau des coefficients de TCHUPROV (cf. § 1.1.2 et tableau 1) un arbre valué ou seules sont représentées les liaisons relativement les plus importantes.



(Les libellés en clair des variables sont donnés au § 1.2.).

On remarque alors que les regroupements a priori des variables (localisation - conditions - description) sont très arbitraires au vu des liaisons rencontrées et que la variable 24 (hauteur angulaire) semble jouer un rôle pivot entre ces différents groupes sans qu'il soit pour autant possible d'avancer une quelconque interprétation en termes de causalité.

Ce schéma illustre seulement la complexité du problème et montre qu'il serait illusoire de vouloir isoler des sous-systèmes "clos" de variables, indépendants des autres paramètres.

	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1	.056	.052	.067	.043	.059	.027	.023	.004	.023	.028	.038	.021	.046	.041	.036	.034	.041	.020	.049	.025	.028	.030	.053	.034	
2		.078	.052	.038	.030	.061	.043	.011	.061	.049	.031	.036	.046	.062	.046	.066	.059	.048	.067	.052	.042	.044	.047	.060	
3			.086	.048	.049	.059	.059	.018	.046	.038	.052	.040	.034	.059	.055	.029	.108	.054	.042	.033	.029	.046	.053	.049	
4				.048	.034	.047	.038	.010	.039	.024	.051	.040	.030	.055	.043	.024	.064	.074	.048	.033	.030	.031	.062	.081	
5					.034	.054	.024	.032	.027	.026	.061	.127	.031	.048	.043	.056	.038	.068	.039	.054	.027	.043	.128	.050	
6						.039	.017	.000	.029	.081	.030	.021	.011	.050	.033	.014	.041	.031	.032	.041	.014	.019	.029	.029	
7							.090	.036	.049	.063	.037	.065	.056	.080	.051	.051	.064	.045	.045	.038	.038	.043	.080	.059	
8								.019	.030	.041	.050	.014	.019	.069	.058	.010	.046	.052	.030	.024	.021	.047	.050	.040	
9									.017	.036	.030	.019	.010	.016	.007	.005	.037	.037	.014	.009	.016	.006	.020	.036	
10										.055	.031	.045	.026	.047	.040	.034	.042	.035	.045	.030	.033	.029	.060	.049	
11											.057	.045	.047	.052	.045	.023	.068	.027	.036	.041	.031	.076	.061	.038	
12												.046	.028	.044	.033	.019	.079	.053	.043	.025	.053	.043	.102	.047	
13													.019	.055	.048	.031	.065	.046	.027	.050	.027	.017	.194	.041	
14														.046	.044	.044	.039	.020	.019	.022	.009	.015	.032	.033	
15															.084	.038	.072	.059	.069	.064	.058	.057	.070	.070	
16																.049	.080	.049	.036	.019	.024	.049	.044	.057	
17																	.047	.029	.050	.034	.023	.025	.038	.038	
18																		.048	.052	.056	.051	.055	.063	.058	
19																			.075	.043	.030	.054	.086	.064	
20																				.082	.102	.047	.073	.050	
21																					.149	.028	.061	.035	
22																						.019	.040	.048	
23																							.046	.024	
24																								.059	

TABLEAU N° 1 :

TABLEAU DES COEFFICIENTS DE TCHUPROV

Si $P_{k\ell}$ désigne la probabilité pour que les modalités k et ℓ et 1 des variables i et j soient prises :

$$T_{ij}^2 = \sum_{k,\ell} \frac{(P_{k\ell} - P_k \cdot P_\ell)^2}{P_k \cdot P_\ell} / \sqrt{m_i - 1} \cdot \sqrt{m_j - 1}$$

m_i : nombre de modalités que prend la variable i .

$$P_k = \sum_{\ell} P_{k\ell}$$

$$P_\ell = \sum_k P_{k\ell}$$

2.2. DESCRIPTION DES MODALITES

Mais, comme les natures des variables sont très hétérogènes et leur nombre important (25), il est préférable, dans un premier temps, d'analyser chacun des groupes séparément avant de chercher à expliciter les liaisons inter-groupes globalement ou plus particulièrement pour certains couples de variables.

Pour chacun de ces groupes, les figures suivantes représentent les projections, sur les plans définis par les deux premiers axes factoriels, des "vecteurs" modalités. Le pourcentage entre parenthèse est la "part d'inertie" expliquée par ce plan.

2.2.1. Localisation (figure 1)

Peu de faits marquants sont notables dans ce groupe dont les variables les plus liées (cf. tableau 1) sont l'estimation de l'heure et la région sans qu'il soit possible d'en tirer une signification. On remarque seulement qu'en mars 1974, dans une période favorable aux confusions*, un nombre relativement important d'observations matinales ont été relatées dans le sud-ouest.

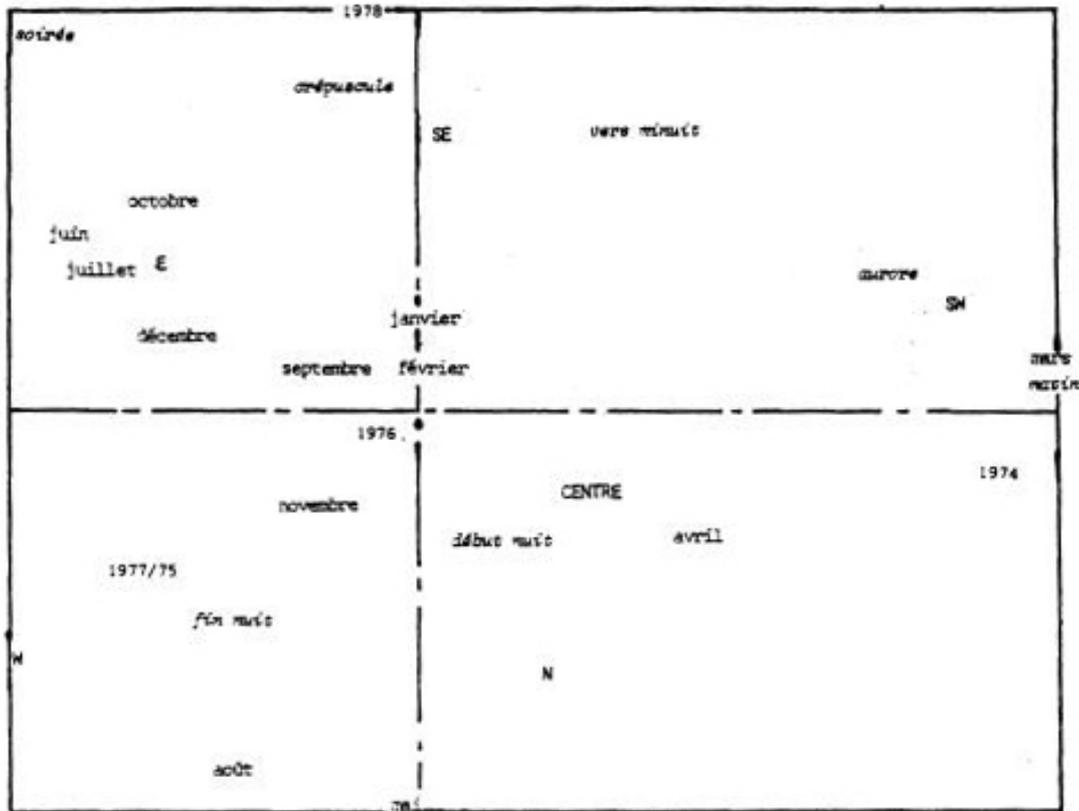
Voir la figure 1.

(*) Le pic de mars 74 a déjà été signalé (cf. BES-80) et suit directement une série d'émissions radio sur le problème "OVNI".

2.2.2. Conditions sociologiques (figure 2)

L'étude de ce groupe de variables, dont les plus liées sont la profession et l'âge des témoins, ne réserve aucune surprise. Le premier axe prend en compte cette liaison et discrimine nettement les groupes armée-police et non-actifs.

FIGURE 1 (13.20)



En effet, une proportion importante de femmes, d'adolescents ou de personnes âgées est non-active alors qu'évidemment le groupe armée-police recrute essentiellement des adultes de sexe masculin.

Le deuxième axe représente le nombre de témoins. Il oppose les observations faites dans une zone isolée ou déserte par un témoin unique, à celles faites dans les hameaux ou petits villages par plusieurs témoins (≥ 4). Les agriculteurs se trouvent plus généralement dans la première catégorie alors que les patrons, cadres supérieurs ou titulaires d'une profession libérale sont des témoins principaux dans la deuxième. Dans un troisième cas, on remarque que, parmi les témoins "employés", une majorité font leur observation en banlieue ou dans une ville conjointement à un autre témoin.

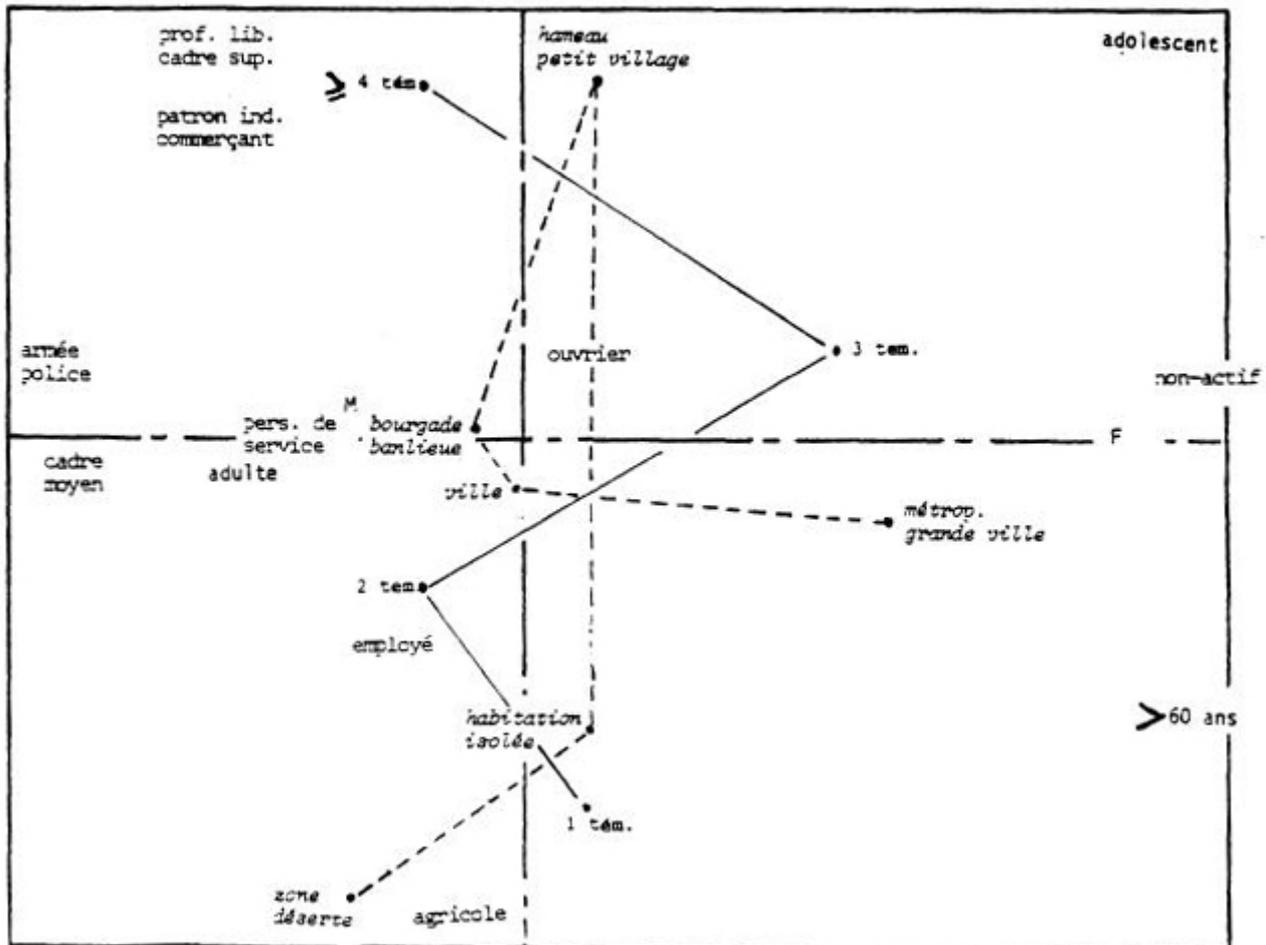
Il n'est guère possible de pousser l'analyse au-delà de ces évidences. La population étudiée ne semble donc pas présenter de particularité sociologique à ce niveau élémentaire*. La liaison qui apparaît entre la nature du lieu d'observation et le nombre de témoins** ne présente pas d'interprétation évidente. Le nombre relativement important d'observations multiples (> 4), faites dans un hameau ou une banlieue par rapport au nombre de celles faites en ville, montre que le nombre de témoins n'est pas directement proportionnel à la densité de population. Il serait abusif d'essayer d'interpréter cette "tendance" compte tenu de la complexité des facteurs mis en jeu et du peu de données.

Voir la figure 2.

(*) Sans oublier pour autant les sous ou sur représentations (déjà relevées en BES-80) de certaines classes par rapport à la population totale et probablement introduites par le filtre que constitue la saisie des informations.

(**) Le test du X^2 conduit à rejeter l'hypothèse d'indépendance au seuil .05

FIGURE 2 (17.04)

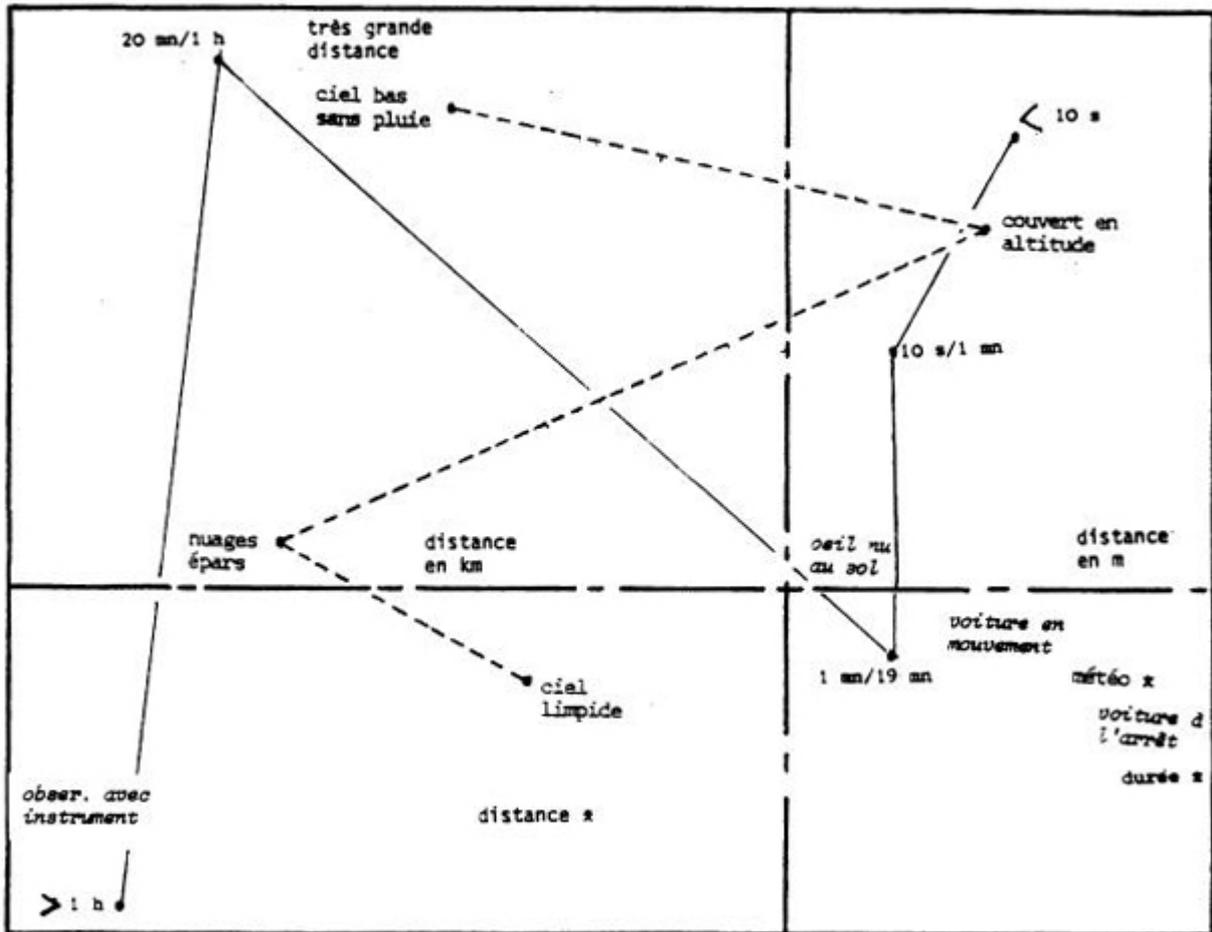


2.2.3. Conditions physiques (figure 3)

Ce groupe, malgré le manque de finesse du codage des modalités (surtout pour l'estimation de la distance^{***}) amène des remarques plus originales.

(^{***}) La finesse d'un codage est fonction de la précision des témoignages mais aussi, pour éviter les classes peu fournies, du nombre des obs.

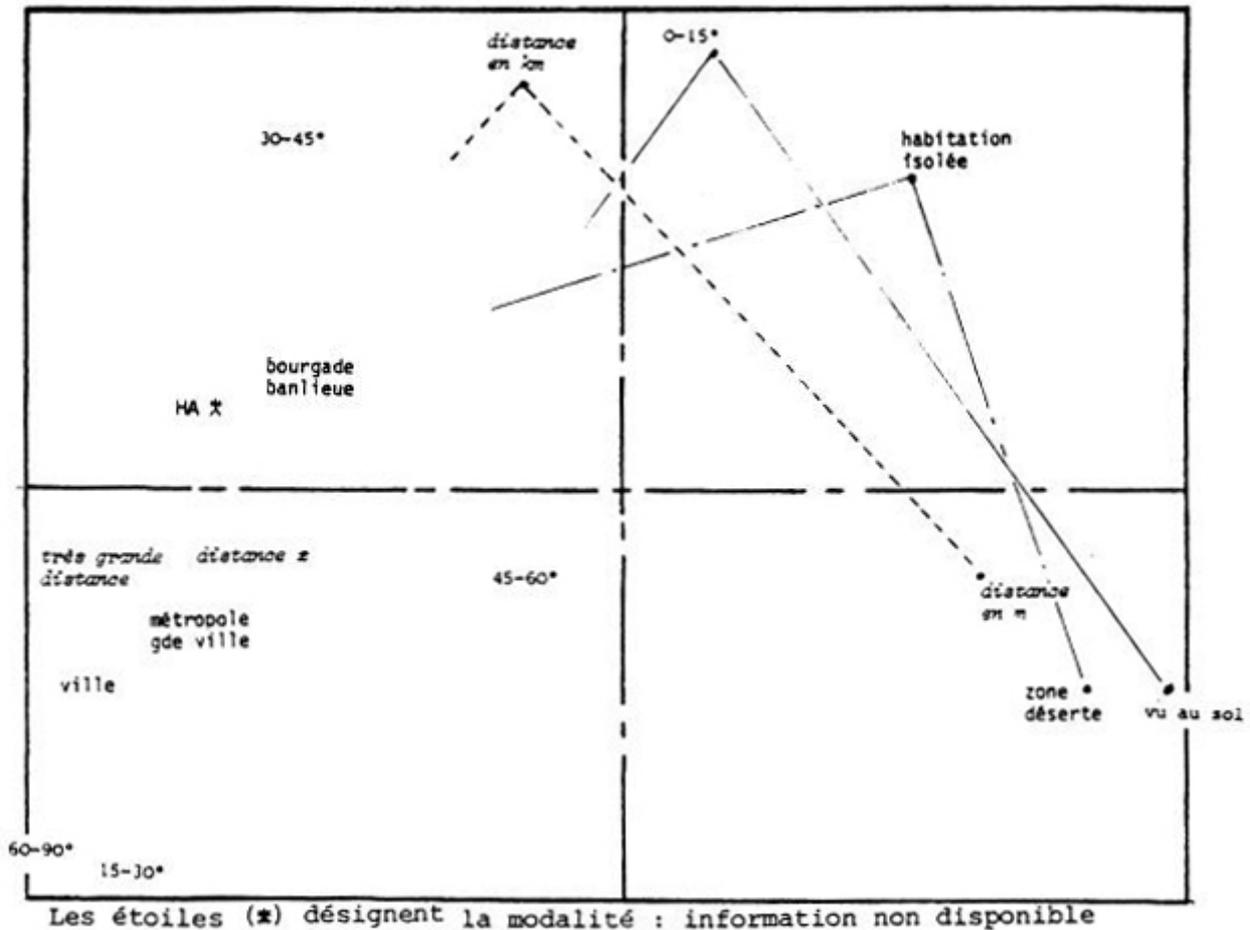
FIGURE 3 (21.35)



Les variables sont relativement peu liées et comme la variable hauteur angulaire est comparativement très liée à l'estimation de la distance, une autre analyse (fig. 4)**** comprenant également la nature du lieu a été effectuée.

(****) Ceci anticipe sur le § suivant (liaisons inter-groupes) mais comme certaines variables (durée, distance, hauteur angulaire,...) concernent à la fois le phénomène et les conditions dans lesquelles il a été observé, il semble artificiel de conserver une frontière trop stricte entre ces deux groupes.

FIGURE 4 (23.13)



Dans la première analyse, (figure 3), le premier axe discrimine fortement les observations brèves (- de 20 mn) de celles plus longues (+ de 20 mn). Du côté des observations brèves se trouve un regroupement facilement interprétable entre les observations faites dans une voiture en marche ou à l'arrêt et les témoignages moins bien documentés (pas d'information sur la météo ou sur la distance).

De l'autre côté, on remarque que les observations, dont la distance estimée est très grande, durent relativement longtemps (20 mn/h) alors que l'emploi d'un instrument (jumelles, appareil photo, ...) se rencontre plus particulièrement pour des observations longues (> 1 heure).

Le deuxième axe représente entre autres, la nébulosité croissante et si l'on distingue deux classes d'observations :

- à droite de l'axe, celles où la distance estimée est faible,
- à gauche, celles où la distance est importante.

On remarque que, dans chacune des classes prises séparément, la durée décroît avec la nébulosité.

L'estimation de la distance est un paramètre fondamental dont les caractéristiques apparaissent peu dans cette analyse ; il est nécessaire de lui réserver un traitement plus particulier.

Voir figure 3 (page précédente).

2.2.4. Estimation de la distance

Les tableaux ci-dessous représentent les "profils horizontaux"* des tableaux de fréquences obtenus en croisant deux variables :

(2) \ (1)	inconnu	au sol	0-15°	15-30°	30-45°	45-60°	60-90°
Inconnue	<u>32.4</u>	12.7	15.5	14.1	8.5	4.2	12.7
en mètres	7.1	<u>54.5</u>	15.2	5.1	9.1	4.0	5.1
en km	18.2	9.1	<u>40.9</u>	0	22.7	9.1	0
très grande	13.3	0	13.3	<u>26.7</u>	<u>20.0</u>	6.7	<u>20.0</u>
Répartition de la variable hauteur angulaire	17.4	31.4	17.9	9.2	11.1	4.8	8.2

TABLEAU 2 : Distance / hauteur angulaire

(1) Hauteur angulaire

(2) Estimation de la distance

(*) Le total de chaque ligne est 100, la dernière ligne représente à titre comparatif la répartition (toujours en %) des effectifs dans les classes de la variable hauteur angulaire.

Le test du X^2 calculé en regroupant certaines modalités (15-45° et 45-90°) conduit à rejeter l'hypothèse d'indépendance au seuil 0.001 (X^2 calculé = 59,07 avec 8 degrés de liberté).

Les valeurs comparativement les plus significatives se retrouvent sur la diagonale exprimant ainsi la tendance du paramètre "estimation de la distance" à croître avec la hauteur angulaire, site de l'observation.

Deux hypothèses peuvent alors être avancées mais sans qu'il soit encore possible de trancher :

- 1) en supposant que la distance est correctement estimée et que le phénomène étudié est équi-réparti dans l'espace, il serait logique, en considérant que le champ de vision d'un observateur croît avec la hauteur angulaire, d'obtenir de tels résultats (voir aussi Note Technique GEPAN N°1 Chapitre 6, paragraphe 4.1. : discussion de la loi de Bouguer) ;
- 2) l'estimation de la distance obéit à un mécanisme perceptif en étroite relation avec le cadre de référence dans lequel est perçu le phénomène (cf. Jim-81). Ainsi, par exemple, la distance estimée serait celle du point de référence angulairement le plus proche alors que le phénomène observé pourrait en être, en réalité, à une toute autre distance.

Ces hypothèses* ne sont pas nécessairement exclusives. Il s'agit avant tout de tendances statistiques, à affirmer ou infirmer par d'autres corpus d'observations, et non de lois strictes.

(*) La première hypothèse postule implicitement l'existence d'un phénomène alors que la deuxième ne suppose que la bonne foi du témoin qui décrit sa perception.

Le tableau suivant suggère des remarques allant dans le même sens (voir tableau n° 3 - Page suivante).

La distance est souvent inconnue en même temps que les conditions météo ou lorsque le ciel est limpide, son estimation en m ne semble pas dépendre de la nébulosité alors qu'une estimation en km nécessiterait un ciel limpide. Le dernier point semble contradictoire puisque c'est surtout par temps plus ou moins couvert que la distance est jugée trop grande pour être précisément estimée.

Ces remarques peuvent être interprétées dans le sens de la deuxième hypothèse : c'est la présence des nuages qui permet aux témoins d'estimer très grande la distance. Ainsi celle-ci est très souvent non estimée lorsque la météo n'est pas mentionnée ou lorsque le ciel est limpide.

TABLEAU 3 : Distance . nébulosité*

(2)\ (1)	inconnue	Ciel limpide	Nuages épars	couvert en alt.	ciel bas sans pluie
inconnue	<u>43.7</u>	40.8	5.6	7.0	2.8
en mètres	41.4	35.4	5.1	10.1	8.1
en km	27.3	<u>45.5</u>	4.5	13.6	9.1
très grande	13.3	33.3	<u>13.3</u>	<u>20.0</u>	<u>20.0</u>
Répartition de la variable nébulosité	38.6	38.2	5.8	10.1	7.2

(1) Nébulosité

(2) Estimation de la distance

(*) L'hypothèse d'indépendance est rejetée au seuil .05 (X^2 , calculé = 9.53 avec 4 degrés de liberté).

TABLEAU 4 : Distance. nature du lieu**

(2)\ (1)	Désert, mer, etc.	Habit. isolée	Hameau p. village	Bourgade Banlieue	Ville	Métropole Gde ville
Inconnue	8.6	15.7	18.6	<u>32.9</u>	15.7	8.6
en mètres	<u>12.2</u>	<u>24.5</u>	<u>30.6</u>	16.3	12.2	4.1
en km	4.3	8.7	<u>39.1</u>	<u>30.4</u>	8.7	8.7
très grande	0	0	0	<u>43.8</u>	<u>50.0</u>	6.3
Répartition de la variable nature du lieu	9.2	17.9	25.1	25.6	15.9	6.3

(1) Nature du lieu

(2) Estimation de la distance

(**) L'hypothèse d'indépendance est rejetée au seuil .01 (X^2 calculé = 22.06 à 6 degrés de liberté)

Les estimations exprimées en mètres ou en kilomètres se feraient alors, en général, par rapport à l'environnement terrestre (faible hauteur angulaire du site du phénomène) sans tenir compte de la nébulosité.

Ceci serait cohérent avec les résultats du tableau 3 mais les effectifs sont trop faibles pour obtenir des résultats significatifs en croisant les trois variables.

A l'exception des cas d'observations dans les grandes villes ou métropoles (dans 46 % de ces cas la distance est inconnue), le tableau 4 semble cohérent avec celui croisant distance estimée et hauteur angulaire ; l'estimation de la distance croît, avec les possibilités de vision en site qui dépend raisonnablement de la densité d'habitat.

Ceci est confirmé par l'analyse factorielle (figure 4) des trois variables : distance, hauteur angulaire et nature du lieu, qui fait apparaître trois groupes de modalités :

- zone déserte, habitation isolée
distance en mètres
vu au sol
- hameau, petit village
distance en km
0 à 15°
- reste des modalités.

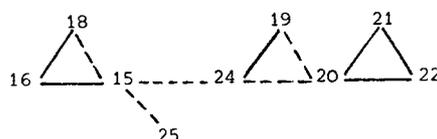
Ne connaissant pas encore quelle a pu être l'influence de l'expertise des cas, il est inutile de poursuivre plus en détails l'étude de l'estimation de la distance qui sera poursuivie sur le fichier complet puisque l'expertise ne fait que filtrer les témoignages sans, bien sûr, modifier leur contenu.

Il est de plus notable que, parmi la faible population de cas d'observations très éloignés (7,7 % beaucoup ont dû être éliminés à l'expertise comme confusion probable avec un phénomène astronomique), aucun ne soit relevé dans une zone à faible densité de population. Ceci invite à rechercher l'influence éventuelle de facteurs anxiogènes (témoin et/ou lieu isolé, frayeur ressentie...) sur l'estimation que fait le témoin de la distance du phénomène.

2.2.5. Description de l'observation

a) Préliminaire

Ce groupe comprend un nombre de variables (12) et de modalités (83) relativement important en comparaison du nombre de témoignages. Mais, de l'analyse du § 1.1.2. sur les liaisons entre variables, émerge une certaine structuration, rappelée ci-dessous, des variables de ce groupe.



Il apparait trois sous-groupes :

- description du comportement :
 - 19 : caractère principal de la trajectoire
 - 20 : caractère secondaire de la trajectoire
 - 21 : estimation de la vitesse
 - 22 : estimation de l'accélération
 - 24 : hauteur angulaire
- description du phénomène:
 - 15 : forme de l'objet principal
 - 16 : luminosité principale
 - 18 : couleur
- variables très peu liées :
 - 14 : nombre d'objets
 - 17 : luminosité secondaire
 - 23 : bruit
 - 25 : azimut

incitant ainsi à dichotomiser l'ensemble des variables pour l'étude de leurs modalités. La liaison marquée entre caractère principal de la trajectoire et hauteur angulaire n'est pas significative puisqu'elle provient d'une redondance du codage. La variable 19 contient en effet les modalités "atterrissages" et "station près du sol" calquant les modalités "vu au sol" et "0-15°" de la variable 24.

Pour éliminer cette liaison parasite, cette dernière variable est rattachée à celles du groupe "conditions physiques".

D'une façon similaire, en étudiant le tableau ci-dessous :

(2)\(1)	inconnue	Lent Immobile	Rapide	Variable	Fulgurante
Inconnue	15.6	59.6	11.0	11.0	2.8
Faible	7.1	78.6	0	14.3	0
Variable	0	18.8	6.3	75.0	0
Élevée	4.5	13.6	40.9	9.1	31.8
Répartition de la variable vitesse	11.8	50.9	13.7	17.4	6.2

- (1) Estimation de la vitesse
- (2) Estimation de l'accélération

TABLEAU 5

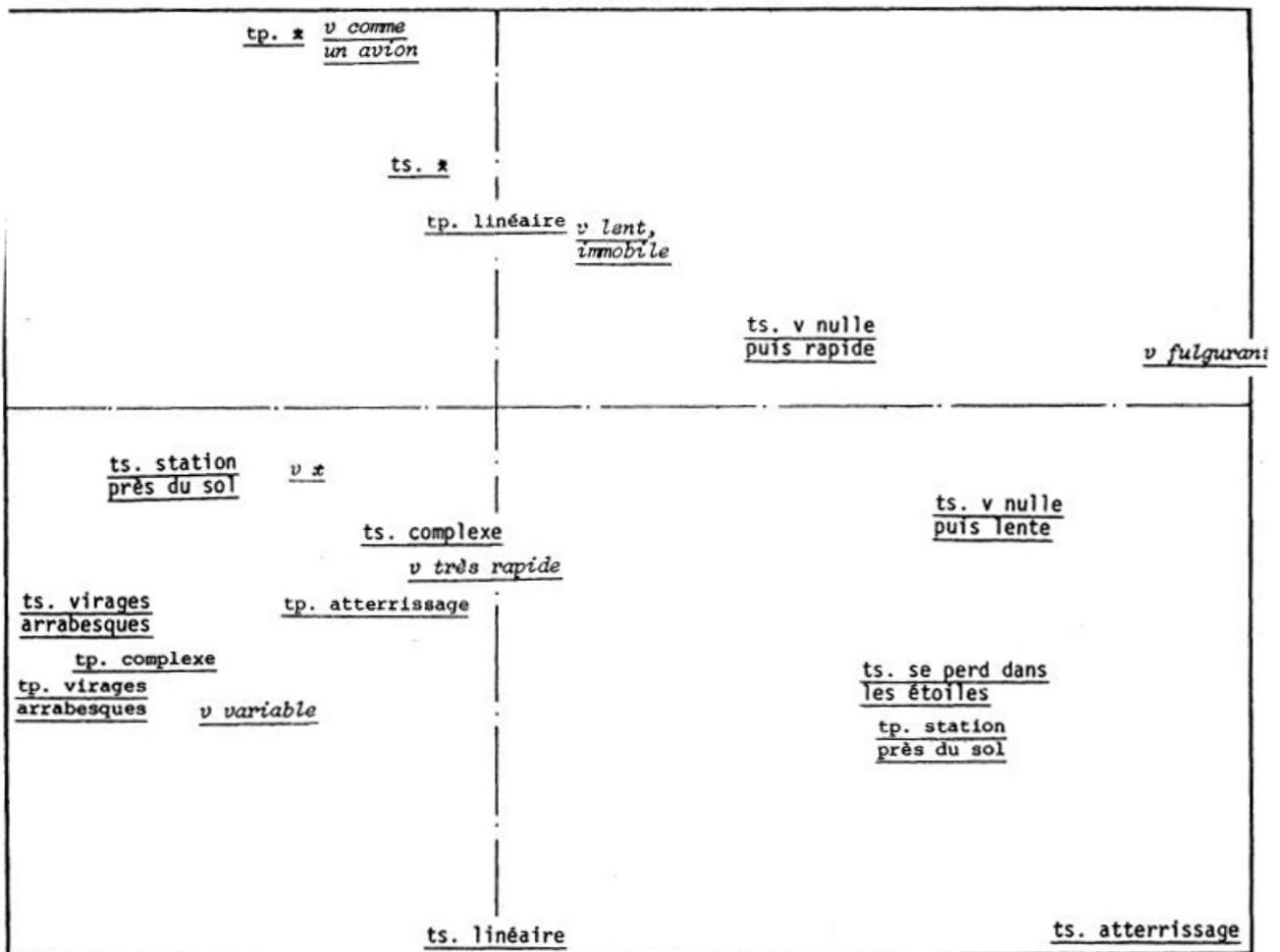
On remarque que l'estimation de l'accélération qui n'est exprimée que dans 33 % des cas, est très souvent une copie de l'estimation de la vitesse. Ne disposant pas d'information supplémentaire à ce sujet, l'accélération est un paramètre difficilement exploitable car, délicat à estimer, il semble faire l'objet d'une confusion* avec la vitesse dans la description des témoins ; Porteur de trop peu "d'information", il est momentanément écarté.

(*) Remarque déjà développée en (Mar-80) sur les mêmes données.

b) Comportement du phénomène (figure 5)

Les axes factoriels n'admettent pas d'interprétations évidentes, on remarque seulement sur cette figure des directions privilégiées de dispersion des modalités dont les plus significatives (pour interpréter une direction) sont celles qui admettent les représentations les plus éloignées de l'origine. En effet les modalités proches de l'origine, formant un angle important avec le plan de projection, sont donc mal représentées et les proximités relevées peuvent n'être que fortuites et sans signification.

FIGURE 5 (17.5)



D'autre part, on remarque que les variables tp (caractère principal de la trajectoire) et ts (caractère secondaire de la trajectoire) comportent des modalités identiques rendant la tâche d'autant plus délicate. Ceci provoque en particulier un éparpillement des modalités : près du sol - atterrissages et une opposition sur le deuxième axe entre : trajectoire linéaire principale et secondaire*.

(*) L'évocation de ces différents problèmes est reprise au § 4.2.

En définitive, nous nous limiterons à relever trois tendances de comportement type des phénomènes décrits :

- a) vitesse comparable à celle d'un avion ou lent, immobile, trajectoire non décrite ou linéaire, sans précision supplémentaire ;
- b) trajectoire complexe, virages, arabesques et/ou linéaire, vitesse variable ;
- c) atterrissage ou station près du sol, vitesse nulle puis lente ou fulgurante, se perd dans les étoiles.

tout en sachant que la plupart des combinaisons sont possibles.

Voir figures 5 (page précédente).

c) Description du phénomène (figure 6)

Les remarques précédentes s'appliquent aussi à ce groupe de variables (redondances entre caractères secondaires et principaux de la luminosité) quoique cette fois un début d'interprétation peut expliquer le premier axe. En effet, celui-ci discrimine les phénomènes non-lumineux (à droite) des phénomènes lumineux. Nous relèverons essentiellement 4 directions de dispersions représentant les tendances les plus marquées :

- 1) phénomène en forme de canotier ou couronner non lumineux ou dont la luminosité n'est pas précisée, pouvant présenter des faisceaux. La couleur est grise ou autre que celles choisies pour le codage, et le phénomène peut s'accompagner d'un bruit plutôt aigu (sifflement, bruit de gaz comprimé ou de moteur électrique) ;
- 2) de forme cylindrique, parfois trois objets de luminosité intense, de couleur rouge ou non précisée, accompagné d'un son plutôt grave (bourdonnement, explosion) ;
- 3) de forme sphérique, couleur blanche ou changeante, luminosité, luminosité brillante, peut-être variable non périodique ;
- 4) la forme n'est pas précisée, 1 ou 4 objets de couleur jaune, luminosité intense, peut-être accompagné d'un bruit non caractérisé.

FIGURE 6 (10.19)

<p>br. plutôt grave</p> <p>cl. rouge</p> <p>3 obj.</p> <p>cl. x</p> <p>cl. blanc</p> <p>fr. sphérique</p> <p>br. silence</p>	<p>ép. intense</p> <p>fr. cigare cylindre</p> <p>ls. autre</p> <p>fr. conique toupie</p> <p>fr. croissant</p> <p>ls. clignotant</p> <p>1 obj.</p> <p>ls. x</p> <p>cl. plusieurs</p> <p>br. plutôt aigu</p> <p>fr. canotier couronne</p> <p>ép. non lumineux avec faisceaux</p> <p>cl. gris</p> <p>ép. lumineux</p> <p>cl. autre</p>
<p>ép. brillant</p> <p>ls. variable non périodique</p> <p>cl. changeante</p> <p>br. x</p> <p>ép. faible</p> <p>2 obj.</p> <p>ls. + ou lumineux</p> <p>cl. jaune</p> <p>4 obj. et +</p> <p>ls. intense</p> <p>fr. x</p> <p>br. autre</p>	<p>fr. autre</p> <p>cl. orange</p> <p>fr. oeuf, ovale ovoïde</p> <p>ép. autre</p> <p>fr. disque soucoupe</p> <p>fr. carrée parallélépipède</p> <p>ép. x</p> <p>ép. non lumineux</p>

d) Description conjointe du phénomène et de son comportement

Il s'agit de vérifier si, parmi les différents types de tendances (comportementales ou descriptives) décrites aux paragraphes précédents, certaines se trouvent conjointement renforcées dans l'analyse réunissant les deux groupes de variables (voir figure 7 page suivante).

Le résultat est plutôt négatif. Il paraît difficile d'associer un comportement à une description type de phénomène. Nous ne pouvons guère que très sommairement couper en deux le plan de projection par la deuxième diagonale qui discrimine donc :

- le comportement c associé au groupe 1 ainsi qu'à un nouveau groupe (5. couleur orange, croissant disque ou soucoupe, leur faible) ;
- des comportements a et b associés aux groupes 2, 3 et 4.

Le résumé de ces regroupements empiriques des modalités est présenté dans le tableau 6 (voir page suivante).

De même qu'ont été comparés les types de tendances (descriptives et comportementales) il s'agit de confronter les regroupements de modalités obtenus aux informations apportées

par les autres groupes de variables (localisation, conditions d'observation). Ces dernières prennent alors le statut de variables "explicatives" dans le sens où elles pourraient "expliquer" (cf. LMT-77) les dispersions rencontrées dans les descriptions des phénomènes ou de leur comportement.

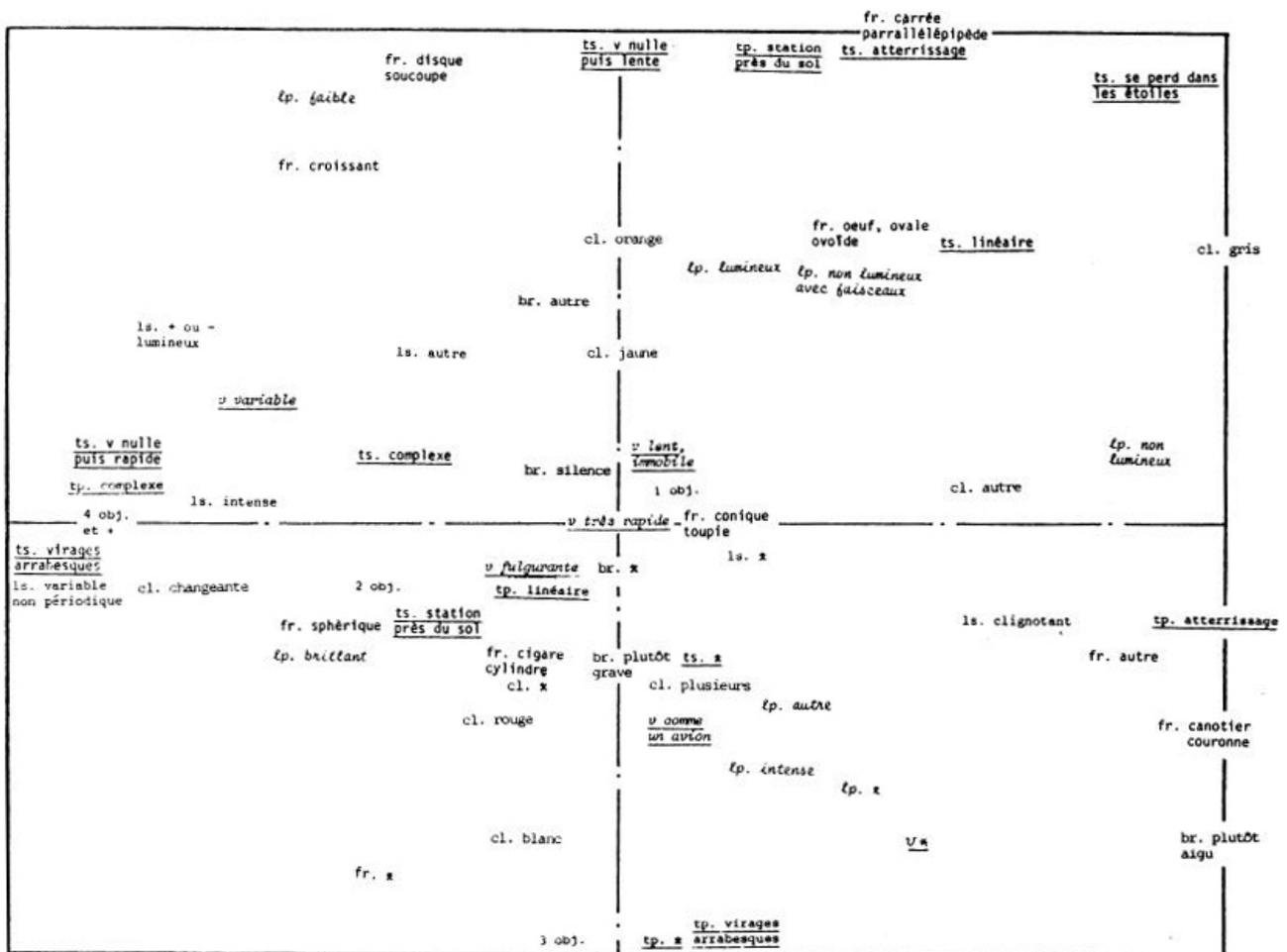
2.2.6. Liaisons inter-groupes

La façon la plus simple de procéder, pour une première approche, consiste à projeter les modalités des variables non considérées pour l'analyse d'un groupe sur les plan factoriels précédemment décrits. On obtient ainsi une représentation des modalités des variables qui "expliquent au mieux" les dispersions rencontrées et les tendances remarquées permettraient d'orienter le travail d'évaluation des liaisons inter-groupes.

Description / condition et localisation de l'observation

On remarque en fait que ces liaisons sont toujours très faibles : les modalités supplémentaires orthogonales au plan de projection se regroupent essentiellement autour de l'origine. Seules quelques-unes se démarquent de l'agrégat central mais sans pour autant signifier qu'il s'agit des caractéristiques dominantes car elles ne concernent souvent que des nombres restreints d'observations.

FIGURE 7 (8.08)



<p>c) Atterrissage ou station près du sol, vitesse nulle puis lente, se perd dans les étoiles.</p>		<p>Forme de canotier ou couronne, non lumineux ou dont la luminosité n'est pas précisée, pouvant présenter des faisceaux. Couleur grise ou autre que celles codables, le phénomène peut-être accompagné d'un bruit plutôt aigu.</p> <p>Lueur faible, de couleur orange, en forme de disque, ou soucoupe.</p>
<p>a) Vitesse comparable à celle d'un avion ou lente, immobile, trajectoire non décrite ou linéaire, sans précision supplémentaire.</p>		<p>Forme cylindrique, parfois 3 objets, luminosité intense, de couleur rouge ou non précisée, accompagné d'un son plutôt grave (bourdonnement, explosion, ...).</p> <p>Forme sphérique, couleur blanche ou changeante, luminosité brillante, peut-être variable, non périodique.</p>
<p>b) Trajectoire complexe, virages, arabesques et/ou linéaire, vitesse variable.</p>		<p>1 ou 4 objets de couleur jaune, luminosité intense, peut-être accompagné d'un bruit non caractérisé.</p>

TABLEAU 6 : Résumé des regroupements des modalités

On trouve ainsi des associations entre :

- Groupe 1 (phénomènes non lumineux,...) et :
 - . observations diurnes
 - . heure, région inconnue
 - . témoins de moins de 13 ans
 - . ciel bas sans pluie
 - . trace physique
- Groupe c (atterrissages,...) et :
 - . crépuscule
 - . cadre moyen
 - . pluie, neige
- Groupe b (trajectoire complexe,...) et :
 - . heure inconnue
 - . mois de mai
 - . ouvrier
- Groupe a (trajectoire linéaire,...) et :
 - . vu d'avion
 - . profession non précisée

De manière identique, on note des rapprochements entre :

- Obs. matinales en mars 74 dans le S-W (cf. § 2.1.1.) et :
 - . couleur bleue
 - . durée < à 10 s
- Obs. brèves faites dans une voiture... (cf. § 2.1.3.) et :
 - . région inconnue
 - . heure inconnue

. vers midi

- Obs. brèves (< 10 s), ... et :
 - . forme de croix
 - . témoin de - de 13 ans
 - . matin

- De 20 à 59 mn, très grande distance, ciel bas sans pluie et :
 - . forme ponctuelle

- Plus de 1 h, obs. avec instrument, distance non estimée et :
 - . 4 "objets", 2 "objets"
 - . non lumineux avec faisceaux
 - . lumière variable non périodique

En définitive, peu de liaisons marquées apparaissent entre chaque groupe de variables et les autres. On peut seulement noter que certaines modalités de faible effectif regroupent des observations présentant des caractéristiques similaires, sans qu'il soit toujours simple de l'interpréter sauf pour certains cas :

- traces physiques dans des cas "d'atterrissages"
- phénomènes non lumineux lors d'observations diurnes
- longue observation à très grande distance d'un phénomène ponctuel.

Ce travail pourrait être complété par l'étude de certains couples de variables (l'étude exhaustive de tous les couples serait fastidieuse) par une analyse classique du tableau des correspondances. Les couples de variables les plus intéressants sur le plan des interprétations, c'est-à-dire ceux correspondant à une hypothèse de travail (estimation de la distance . hauteur angulaire, profession . forme, etc...) ont déjà été et/ou vont être étudiés plus spécifiquement (cf. § 4.4.). Il resterait alors à analyser les couples présentant les plus fortes liaisons dans le tableau 1 du § 2.1. Ce travail qui ne fait pas émerger d'éléments nouveaux n'est pas décrit ici. On signale seulement quelques correspondances :

- obs. diurne - couleur grise ou métallique
- à partir d'une voiture - zone isolée et une nette opposition
- observation à l'aide d'un instrument - zone déserte.

2.3. RÉSUMÉ DE L'ANALYSE DESCRIPTIVE DES VARIABLES ET DE LEURS MODALITÉS

L'étude des liaisons entre les variables montre que celles-ci sont faibles et seules quelques tendances ont pu être dégagées sans qu'il soit possible de parler de structures du nuage des observations.

Ainsi, la localisation spatiale et temporelle des observations présente une dispersion relativement homogène et ne semble pas avoir d'influence déterminante sur les descriptions de celles-ci, sinon qu'évidemment les phénomènes non-lumineux sont observés de préférence le jour.

Les caractéristiques sociologiques des témoins (âge, profession, sexe) ne se démarquent pas de celles de la population totale à l'exception de la relation observée entre le nombre

de témoins et la nature du lieu (équivalente à la densité de population). Ces résultats vont dans le sens d'une banalité sociologique du fait d'être témoin. Ne sachant rien sur la proportion des témoins connus parmi les observateurs et leurs motivations à témoigner, il est difficile d'en donner une interprétation plus précise.

Les liaisons inter-groupes caractéristiques sociologique/description sont plus facilement étudiées. Elles ont déjà été signalées en Leg-79 et concernent plus particulièrement la profession, le nombre de témoins et la forme décrite. La correspondance observée (cf. § 2.2.6) entre des cas "d'atterrissages" et des témoins de moins de 13 ans est ambiguë. Elle suggère au moins deux interprétations :

- ce seraient des observateurs facilement suggestibles,
- parmi les observations relevées par des enfants jugés "peu crédibles", les experts ont surtout retenu les cas les plus "étranges" car "jugés plus intéressants".

Ces relations juste signalées n'ont pas été plus détaillées ici puisqu'elles font l'objet d'une étude spécifique*, centrée sur les caractéristiques psychosociales des témoins et de leurs observations ; elles nécessitent une information plus riche que celle retenue par le codage.

(*) Étude menée dans le cadre d'une convention avec l'Université PARIS V René Descartes.

Parallèlement à un certain nombre de remarques élémentaires concernant les conditions physiques des observations :

- obs. à partir d'une voiture - peu d'information
- obs. brèves - peu d'information
- obs. à très grande distance - durée importante (20 mn/1 h)
- obs. à l'aide d'un instrument - durée très importante (> 1 h)
- durée d'obs. - elle décroît avec la nébulosité

une problématique importante est posée autour de l'estimation de la distance et des liens entrevus avec la nature du lieu, la hauteur angulaire, la nébulosité éventuellement des facteurs anxio-gènes (isolement, nature de l'obs., etc...) .

Deux types d'hypothèses sont alors en présence (cf. § 2.2.4), l'une attribuant les caractéristiques rencontrées au phénomène, l'autre aux mécanismes perceptifs associés à l'estimation de la distance. Il est évidemment trop tôt pour se prononcer. Une étude dans ce sens est en cours.

Elle devra aussi faire intervenir l'estimation de la taille dans les facteurs à prendre en compte.

Les liaisons observées entre les variables décrivant l'observation ont permis de les diviser en deux sous-groupes, l'un décrivant le comportement, l'autre le phénomène. L'analyse de chacun des sous-groupes, ainsi que l'étude globale révèlent, entre les modalités des différentes variables, un certain nombre de correspondances qui sont résumées en une classification empirique des modalités (cf. tableau 6).

Le problème qui se pose alors est de connaître la validité d'une telle classification ; correspond-elle à des types bien définis d'observations, et donc à une classification de celles-ci, ou bien est-elle la résultante de tendances plus ou moins marquées et associées à des séries de co-occurrence des divers facteurs sur le nuage des observations ? Le peu de dispersion dans les plans factoriels (sphéricité du nuage) corrélatif à la faiblesse des

liaisons nous oriente vers la deuxième possibilité, confirmée par ailleurs au paragraphe suivant.

3. CLASSIFICATION AUTOMATIQUE DES OBSERVATIONS

3.1. PRÉLIMINAIRE

La méthode employée brièvement décrite au § 1.1.*, consiste à regrouper les observations les plus "semblables" c'est-à-dire les plus "proches" au sens d'une certaine métrique. C'est donc, avec l'analyse des variables (§ 2.1.) et l'analyse des modalités (§ 2.2.) une troisième façon d'aborder le corpus des données. Aucune de ces méthodes n'est a priori plus performante que les autres, chacune apporte un point de vue complémentaire et les confrontations des résultats permettent, sinon de valider, du moins de fixer des limites aux interprétations proposées.

(*) Pour plus de détails, se reporter par exemple à LMT-79.

Parallèlement aux analyses portant sur la description des phénomènes et de leurs comportements, trois classifications des observations ont été réalisées. Chacune d'elles fournit une partition en 10 classes, les neuf premières (notées de 1 à 9) apparaissent comme les plus stables à travers une série de classifications élémentaires en 3 classes tandis que la dixième (notée A) regroupe les observations marginales qui évoluent d'une classe à l'autre. Toutes les observations sont alors représentées dans les plans factoriels précédemment calculés où chaque cas est symbolisé par le numéro de la classe à laquelle il appartient (figures 8, 10 et 12). Puis, afin de comparer dispersion des modalités et dispersion des classes, ce sont les centres de gravité de chaque classe, qui sont représentés conjointement à l'ensemble des modalités (figures 9, 11 et 13).

3.2. RÉSULTATS

Les résultats obtenus sont peu significatifs, les représentations des classes sont rarement disjointes et même fréquemment imbriquées. De plus, les classes d'observations obtenues correspondent peu aux classifications des modalités élaborées précédemment.

3.2.1. Comportement du phénomène (figure 8)

Le nombre restreint de variables conduit à des classes comparativement plus distinctes que dans les autres cas de figures ci-dessous. On peut associer alors certaines classes (en terme d'inclusion) aux types de comportements décrits au paragraphe 2.2.5.* :

N° classe	Effectif	Comportement type
1	80	a
7	9	b
4	15	c
5	12	
6	10	

3.2.2. Description du phénomène

La même opération a été entreprise malgré le flou accru :

FIGURE 9

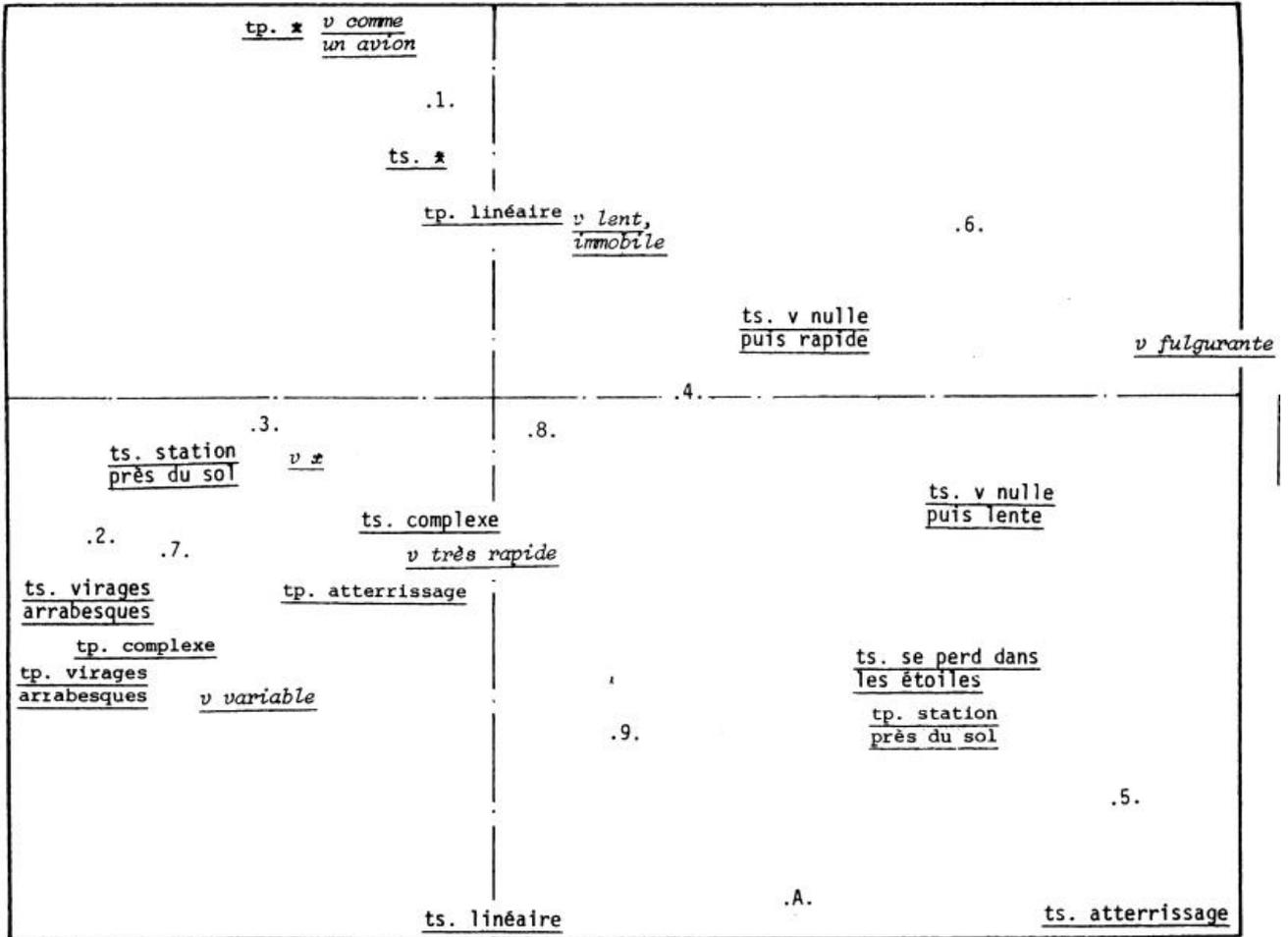


Figure 10

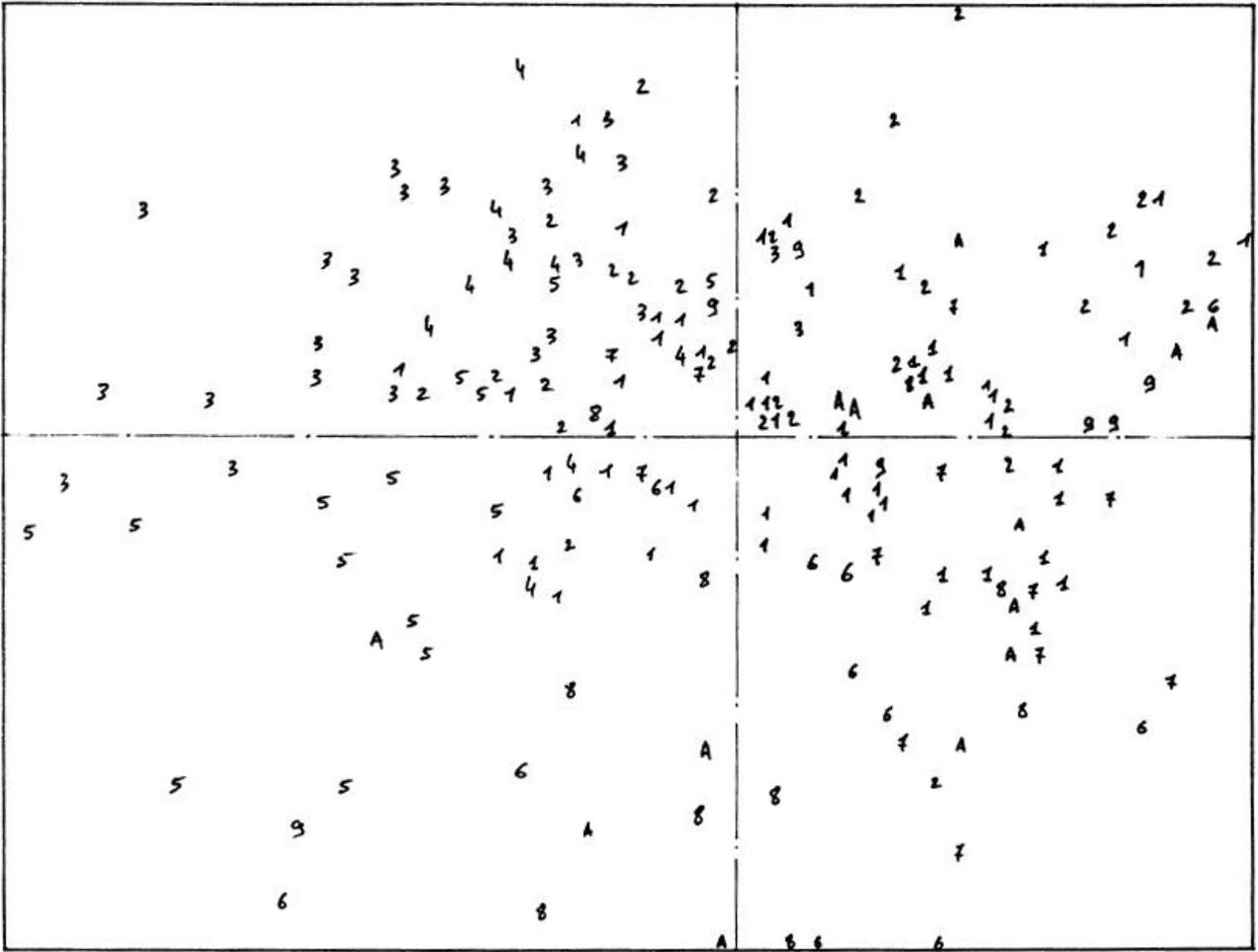
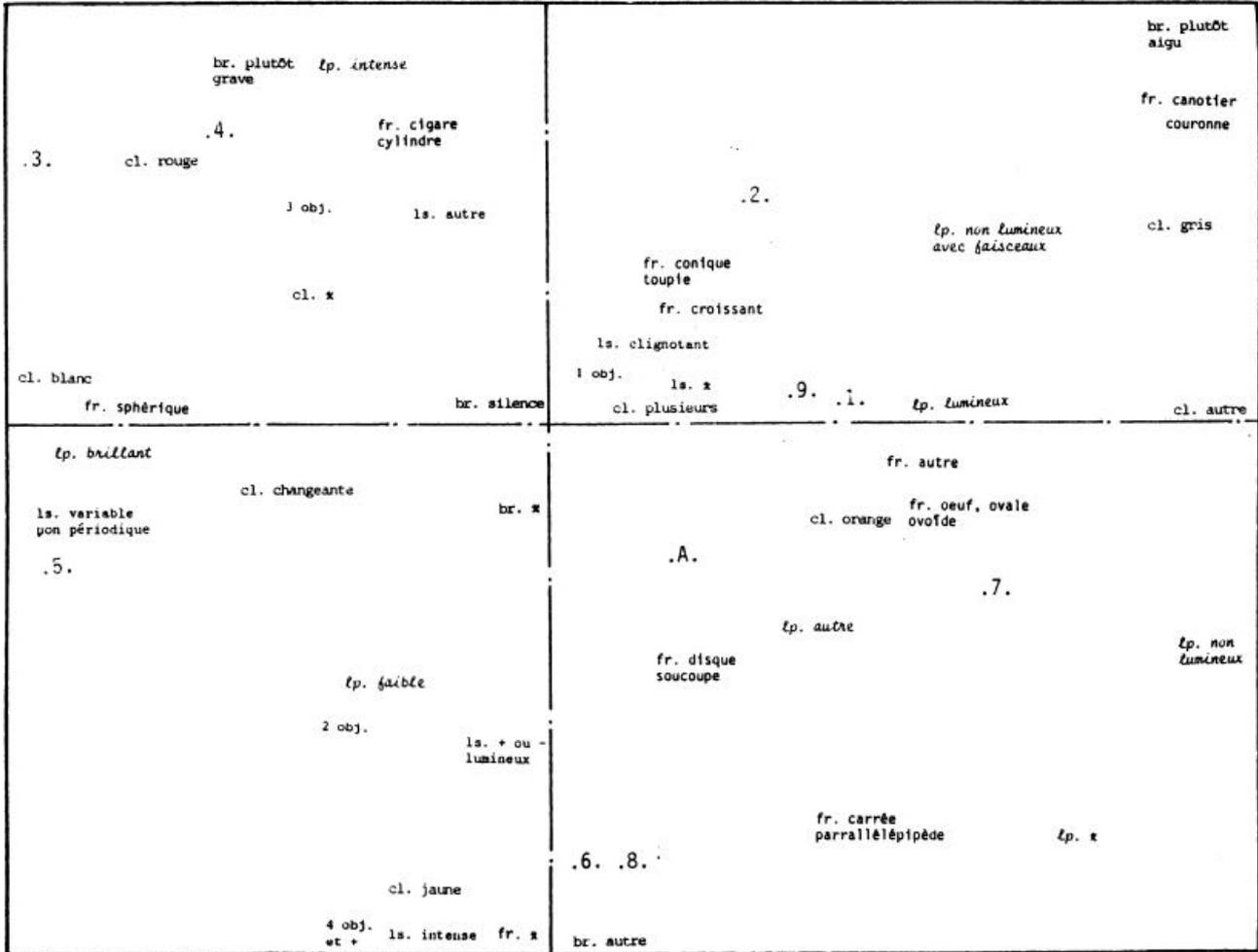


FIGURE 11



3.2.3. Description des phénomènes et de leurs comportements

Il est encore plus difficile de donner un sens à la classification obtenue relativement aux regroupements des modalités. Ainsi, la classe n° 8 qui est comparativement la plus disjointe des autres, se situe de part et d'autre de la deuxième bissectrice utilisée au § 2.2.5. pour partager les types de description.

On se contente alors d'attribuer à chaque classe la liste des modalités qui sont les mieux représentées. (voir tableau 7 page suivante) .

On remarque évidemment que les modalités de plus fort effectif sont le plus souvent cités puisque ce sont elles qui favorisent les proximités entre observations. Les techniques de classification automatique sont donc bien complémentaires des analyses factorielles qui font elles surtout apparaître les modalités de faible effectif moins probables donc plus significatives au sens de la métrique du X^2 .

Dans une situation plus favorable (celle, par exemple, où les deux techniques d'analyses - factorielles et classification - conduiraient à des interprétations se renforçant mutuellement) il serait sans doute intéressant de revenir, après classification, aux procès-verbaux, source des informations.

Les confrontations de certains éléments d'informations pourraient suggérer alors des possibilités d'interprétation ou encore susciter des pistes de recherches en sciences physiques. Mais actuellement, une telle démarche qui sort du cadre strictement statistique de cette étude, ne semble pas suffisamment motivée pour être entreprise.

EFFECTIF	CLASSE	LISTE DES MODALITES LES MIEUX REPRESENTEES
54	1	Rond, sphérique ; brillant intense ; rouge ou blanc ; trajectoire droite ou légèrement courbée ; lent, immobile ; silence.
25	2	Disque, soucoupe ; lumineux ; orange ; trajectoire droite ou légèrement courbée ; lent, immobile ; silence.
15	3	Oeuf, ovale ; lumineux ; orange ; trajectoire droite ou légèrement courbée ; lent, immobile ; silence.
15	4	Cigare, cylindre ; brillant ; plusieurs couleurs ou inconnues ; trajectoire droite ou légèrement courbée ; vitesse variable ; silence.
14	5	Rond, sphérique ; brillant ; variable non périodique ; rouge, orange ou blanc ; trajectoire droite ou complexe ; vitesse nulle puis rapide, silence.
13	6	Oeuf, ovale ; lumineux ; clignotant ; plusieurs couleurs ; station près du sol ; "atterrissage" ; vitesse variable ; silence.
13	7	Rond, sphérique ; brillant ; orange ; trajectoire droite ou complexe ; station près du sol ; lent, immobile ; silence.
9	8	Forme autre ; intense ; plusieurs couleurs ; atterrissage ; vitesse inconnue ; bruit dans les aigus, silence.
9	9	Rond, sphérique ; brillant ; variable non périodique ; orange ou couleurs changeantes ; trajectoire droite ou complexe ; virages, arabesques, vitesse variable ; silence.

TABLEAU 7

Figure 12

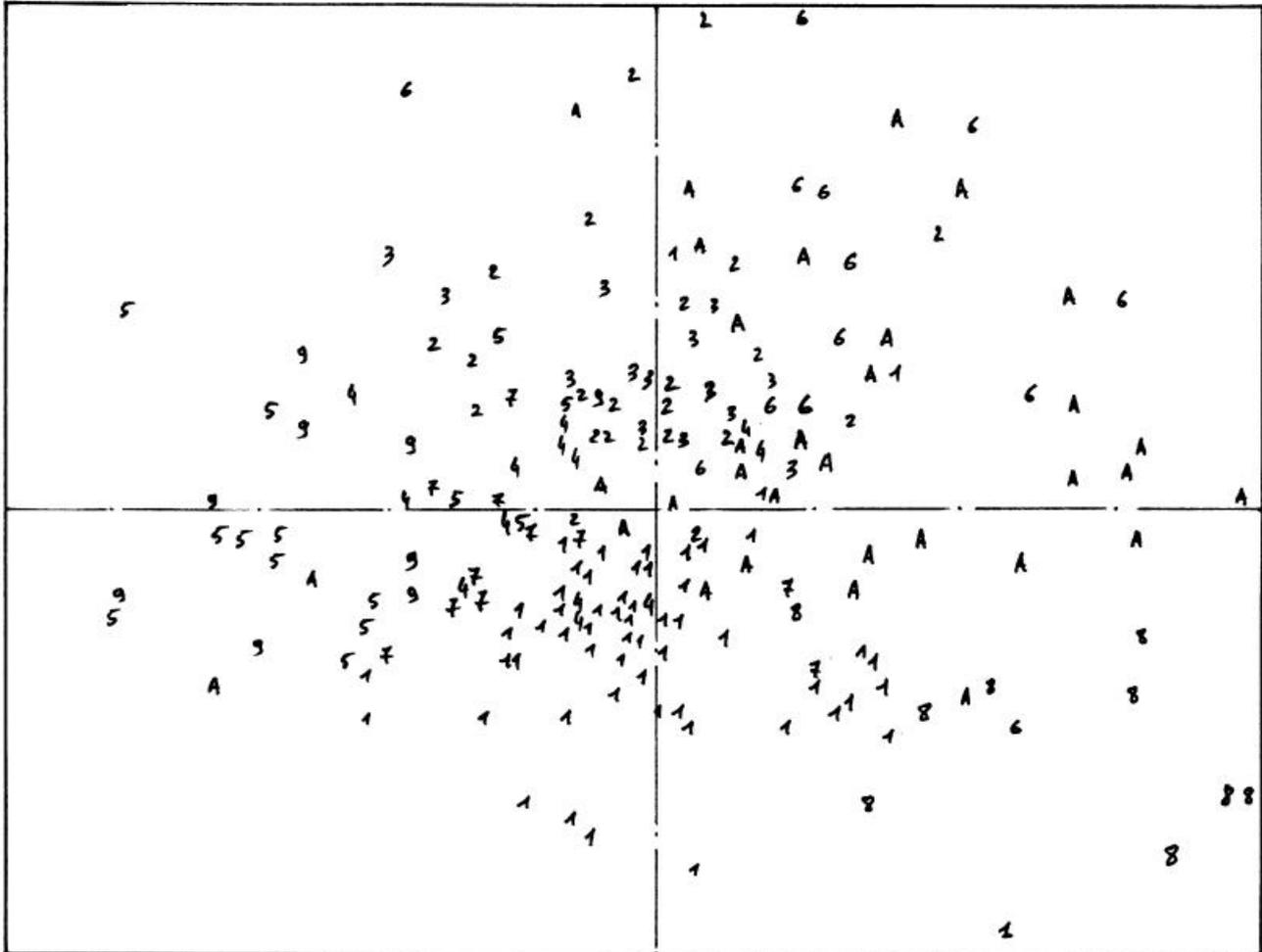
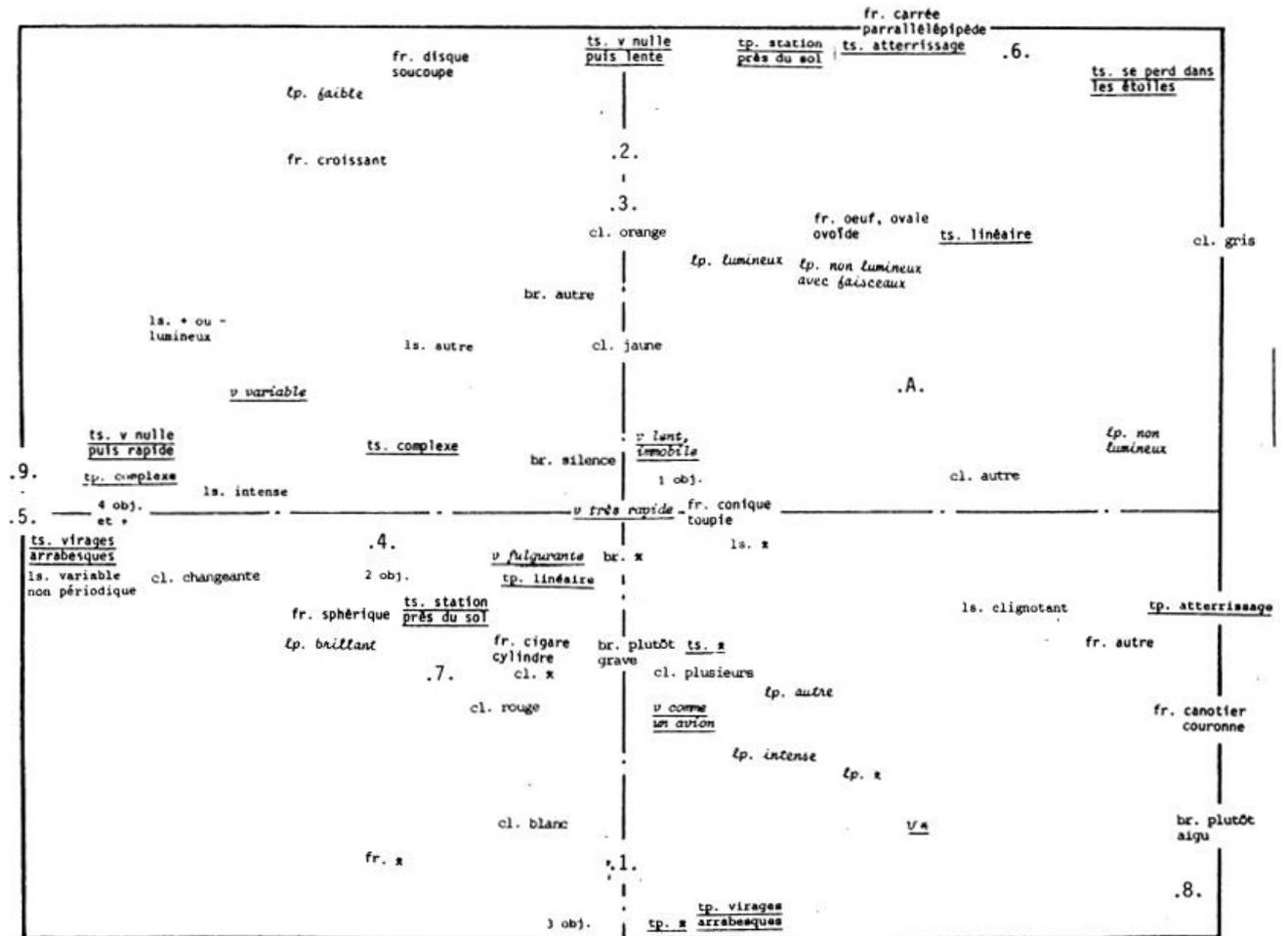


FIGURE 13 :



4. CONCLUSION

4.1. LES TECHNIQUES D'ANALYSES

Les outils employés sont l'objet d'un certain nombre de critiques classiques qu'il ne faut pas perdre de vue surtout lorsque comme dans le cas présent, les faibles dispersions rencontrées (faible "inertie expliquée" des plans factoriels) conduisent à des interprétations dont la validité peut être sujette à caution.

Les analyses factorielles tiennent compte uniquement des "distances" entre les variables prises 2 à 2 et négligent ainsi les effets d'interactions lorsqu'interviennent plus de deux variables. Les techniques existant pour réduire ces effets (cf. BBCP - 77) nécessitent des effectifs importants et n'ont pu être développées. De son côté, la méthode de classification automatique employée (les "nuées dynamiques") est sensible au choix de la métrique utilisée pour mesurer les "proximités" entre observations.

De plus, le modèle mathématique sous-jacent à ces analyses (représentation vectorielle) fait jouer aux modalités des rôles identiques indépendamment des structures (ordre, pré-ordre, ...) que peuvent induire les variables ou des significations qu'elles recouvrent (couleur non signalée, couleur rouge, couleur autre, ...). Ceci fait comme on va le voir sur un exemple ci-dessous, que ces analyses sont particulièrement sensibles aux choix des variables et de leurs modalités et donc aux règles de codage.

4.2. LE CODAGE

Un exemple type de l'influence du codage est donné au 2.2.5. dans l'étude du comportement des phénomènes observés. On remarque en effet que le deuxième axe factoriel de cette analyse (figure 5) discrimine d'une part la modalité "linéaire" de la variable "caractère principal de la trajectoire" et d'autre part cette même modalité "linéaire" de la variable "caractère secondaire de la trajectoire".

Ceci s'explique simplement par le fait qu'on ne code pas deux fois la même information (linéarité) et donc qu'aucune observation ne prend simultanément ces deux modalités. Ainsi comparativement aux effectifs de chacune des modalités, cette co-absence, devenant significative, est prise en compte par l'analyse alors qu'elle n'exprime qu'une insuffisance du codage au lieu d'une caractéristique des témoignages.

Cette redondance se retrouve pour d'autres modalités ("atterrissages", "station près du sol", "0-15°" etc...), accroissant le flou du nuage des observations et les difficultés d'interprétation. Il faut ajouter de plus que le codage ne tient pas compte des différentes phases que peut comporter la description d'un phénomène, ce qui participe encore à la confusion.

4.3. CONCLUSION DE CETTE ETUDE

Compte tenu des critiques qui précèdent, il est clair qu'aucune certitude n'émerge de ce travail, mais pour comprendre ce qu'il peut apporter, il faut le resituer dans le contexte des études statistiques déjà réalisées.

Les mêmes types d'analyses*, effectués sur un autre corpus de données, ont trouvé que les dispersions les plus marquées du nuage des observations provenaient essentiellement

de la variété des sources utilisées pour constituer le fichier (livres ou revues, français ou étrangers...) mais les investigations n'ont guère pu être approfondies.

(*) cf. Der-76, Bor-77 sur le fichier établi par Claude Poher

Un travail plus récent** de comparaison de fichiers a permis de révéler le rôle joué par les différents filtres situés entre les témoignages et l'analyse statistique ; l'étude présente, utilisant des données venant d'une source supposée homogène, et malgré certaines insuffisances du codage, permet de franchir une nouvelle étape comme l'illustre le § 2.2.4, sur l'estimation de la distance, le débat statistique se place maintenant au nœud du problème dans toute la complexité des interactions entre physique et psychosocial, témoignages et environnements telles qu'elles ont déjà été décrites (cf. Bes-80 et ABEJ-81).

(**) cf. Bes-80

En effet, il semble difficile d'être plus précis dans la description et l'interprétation des données recueillies sans que progresse, en parallèle, une connaissance même élémentaire des mécanismes perceptifs et des témoignages dans des conditions psycho-sociologiques expérimentales similaires. C'est une condition nécessaire pour essayer d'évaluer et d'interpréter les influences réciproques du psychologique, du sociologique et du physique ; de discerner, parmi les caractéristiques ou constantes relevées dans les rapports d'observations, celles spécifiques du témoignage humain de celles attribuables aux phénomènes observés dans un débat qui risque de plagier, dans sa forme statistique, le débat classique caractères acquis/caractères innés en génétique***.

(***) Avec tous les pièges statistiques que cela comporte (cf. ChJ-79).

4.4. ...A SUIVRE...

Ainsi, ce travail permet au moins d'orienter les pistes de recherches (psycho-sociologie ; psychologie de la perception) dont le développement est nécessaire à la poursuite d'une étude typologique des phénomènes observés.

Dans cette perspective, le codage actuel, qui s'efforce de réaliser un compromis entre les différentes études possibles, devient insuffisant et chaque recherche doit en développer un, spécifique aux hypothèses à valider. On note par exemple, qu'une étude sur la perception s'attachant aux estimations des distances, des tailles, des vitesses, ..., nécessite un codage homogène et très fin, alors que pour les techniques d'analyse des données et de classification, il faut éviter les modalités redondantes ou à effectif trop faible.

D'autre part, au vu des classifications obtenues (floues ou difficilement interprétables) et de même qu'une étude de la perception portera d'abord sur les rapports d'observations de phénomènes reconnus à posteriori*, il semble intéressant d'orienter l'analyse statistique vers un type de classification dit "avec apprentissage"***. Ceci aura pour effet de considérer les phénomènes non identifiés à leur place, c'est-à-dire en marge des phénomènes, identifiables et identifiés, plutôt que comme une "population statistique" dont l'homogénéité n'est peut-être qu'illusoire***.

Enfin, un point théorique important reste à résoudre : comment les différentes études abordant des domaines complémentaires, vont-elles s'articuler ? Il est, par exemple, évident qu'une connaissance accrue des mécanismes perceptifs doit conduire à préciser l'image statistique du "phénomène OVNI" et, réciproquement, la connaissance des

caractéristiques des phénomènes doit alimenter l'étude de la perception. Mais les outils statistiques existants ne sont pas encore adaptés à une telle "gymnastique".

(*) Observation d'un phénomène dont les caractéristiques physiques sont connues, dans des conditions similaires à la perception d'un phénomène inconnu.

(**) Par analogie avec la télédétection où les parcelles élémentaires du sol (pixels) sont regroupées autour de "points vérité terrain" dont la nature est connue a priori.

(***) Sans oublier qu'on évite ainsi le biais introduit par l'expertise.

BIBLIOGRAPHIE

- ABEJ-81 AUDRERIE D., BESSE P., ESTERLE A., JIMENEZ M.
Note Technique N° 3
CNES/GEPAN - A paraître
- BBCP-77 BAUMERDER A., BONNEFOUS S., CAZES P., PAGES JP.
Codage et analyse du tableaux logiques - Introduction
à la pratique du variables qualitatives
Cahier du B.U.R.O. - n° 27 - 1977 – Paris
- Bes-80 BESSE P.
Étude comparative des résultats statistique élémentaires relatifs aux
observations des phénomènes aérospatiaux non identifiés
CNES/GEPAN 1980 - Note Technique n° 2
- Bor-77 BORDET JP.
Etude statistique multidimensionnelle d'un ensemble d'observations d'OVNI
Réunion du Conseil Scientifique du GEPAN - décembre 77
- Cap-76 CAILLEZ F., PAGES JP.
Introduction à l'analyse des données
SMASH – 1976
- ChJ-79 CHOUCHAN M., JACQUARD A.
A propos d'intelligence et de quotient intellectuel. Rôle d'un "modèle additif" dans
les conceptions d'inné et d'acquis.
Bulletin de l'APMEP – 58° année – n° 321 – dec. 79
- Dap-76 DAUXOIS J., POUSSE A.
Les analyses factorielles en calcul du probabilités et en statistiques : essais
d'étude synthétique
Thèse d'état - Toulouse III - 1976
- Der-76 DERANSART P.
Etude des données du CNES portant sur des observations d'OVNI
Note manuscrite
- JIM-81 JIMENEZ M.
Psychophysique de la perception des distances
CNES/GEPAN - Recherches en cours

- LMT-77 LEBART L., MORINEAU A., TABARD N.
Techniques de la description statistique
Dunod – 1977
- Leg-79 LEGENDRE P.
Rapport d'études statistiques
CNES/GEPAN - Note Technique N° 1 - 1979
- Mar- 80 MARCUS M.
Analyse statistique de cas d'observations d'OVNI
Rapport de D.E.A. - Paris V I – 1980
- Sap-75 SAPORTA G.
Liaison entre plusieurs ensembles de variables et codages de données
qualitatives
Thèse de 3° cycle - Paris VI – 1975
- Sap-78 SAPORTA G.
Théories et méthodes de la statistique
TECHNIP - 1978