

MARYVONNE MERRI

**Aide à la planification dans la résolution de problèmes : conception
de Prosaic, un logiciel d'E.I.A.O. de statistiques**

Publications de l'Institut de recherche mathématiques de Rennes, 1987, fascicule 5
« Séminaires de didactique des mathématiques », , exp. n° 2, p. 1-27

http://www.numdam.org/item?id=PSMIR_1987__5_A2_0

© Département de mathématiques et informatique, université de Rennes,
1987, tous droits réservés.

L'accès aux archives de la série « Publications mathématiques et informatiques de Rennes » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

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

AIDE A LA PLANIFICATION DANS
LA RESOLUTION DE PROBLEMES :
CONCEPTION DE PROSAIC, UN LOGICIEL
D'E.I.A.O. DE STATISTIQUES.

Maryvonne MERRI : Psychologue à l'Ecole Nationale de la Santé Publique.

25 mars 1987

Résumé

Le didacticiel de statistiques descriptives à deux variables qui est présenté ici constitue la maquette du projet d'E.I.A.O. destiné à de futurs cadres de la santé (Ecole Nationale de la Santé Publique - Rennes).

Cette maquette expérimentale bénéficie d'une analyse des notions et des tâches, ainsi que d'une analyse de l'activité de résolution de problèmes. Une représentation hiérarchique sous forme de plans est utilisée.

L'architecture du didacticiel est présentée ainsi que le travail sur la représentation des connaissances sous forme d'objets, d'actions et de démarches, qui a été réalisé.

La caractéristique tutorielle de PROSAIC est la suivante : développer un outil de compréhension des actions de l'élève. Ceci exige l'élaboration d'un modèle à deux composantes : la composante "Résolution du problème par l'élève" et la composante "Compréhension des actions par l'enseignant".

Cette élaboration est la prochaine phase de conception de PROSAIC.

I - CONTEXTE DE LA CONCEPTION DE LA MAQUETTE DE STATISTIQUES.

L'Ecole Nationale de la Santé Publique (E.N.S.P.) élabore depuis 1986 un projet d'Enseignement Assisté par Ordinateur. La première phase de ce projet est la conception puis la réalisation d'une maquette expérimentale.

La forte demande de formation, l'hétérogénéité des cursus et expériences antérieurs des élèves, ainsi que le grand nombre de disciplines abordées ont motivé ce projet.

Ce projet comporte quatre modules d'enseignement :

1. Un tronc commun obligatoire (SOCIO-ECONOMIE DE LA SANTE et DEMOGRAPHIE) ;
2. Un module **METHODES QUANTITATIVES** ;
3. Un module EPIDEMIOLOGIE ;
4. Un module PLANIFICATION.

Le cursus aboutit au module planification pour lequel la mise en place préalable d'outils d'analyse est nécessaire.

Ce projet comporte deux aspects :

- Une description et une gestion de Curricula : les cheminements possibles au sein de l'ensemble des didacticiels sont définis et des profils d'élèves y sont associés ;
- La conception et la réalisation de chaque didacticiel.

Nous nous sommes intéressés pour élaborer une maquette expérimentale à une Unité de Valeur particulière : les **statistiques descriptives à deux variables**.

Les notions qui y correspondent sont replacées dans le schéma 1.

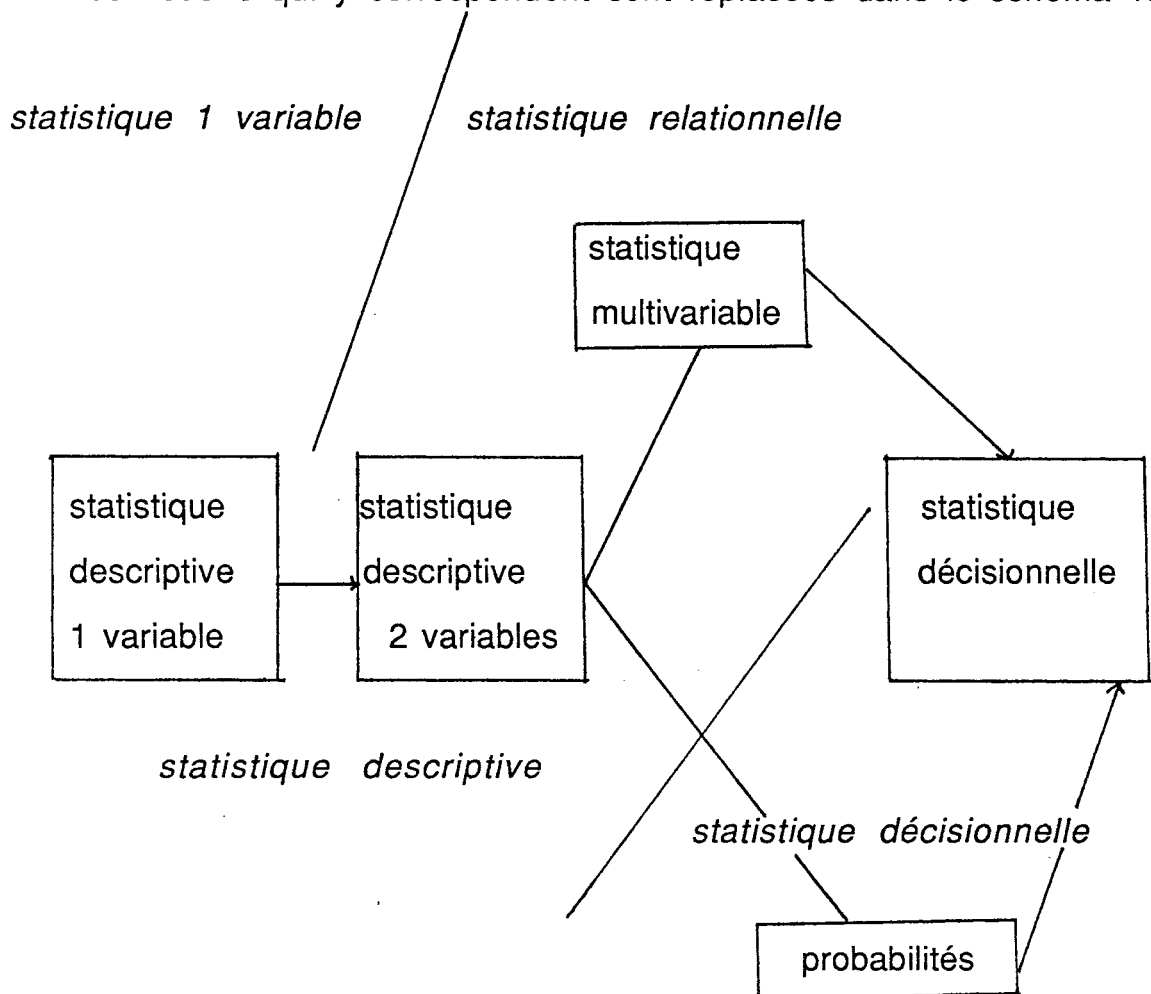


Schéma 1 : Différentes parties des statistiques successivement abordées.

Il apparaît, sur le schéma 1, que l'U.V. "statistiques descriptives à deux variables" est particulièrement intéressante à aborder. En effet,

- Il s'agit de **statistiques descriptives** par opposition aux **statistiques décisionnelles** ;
- Il s'agit de **statistiques relationnelles** par opposition aux **statistiques à une variable**.

II - OBJECTIFS DU DIDACTICIEL

PROBLEME-TYPE : Nous proposons une distribution bi-variée (à deux variables).

En amont, s'est posé un problème de gestion (ici, un directeur d'hôpital veut étudier la durée d'hospitalisation).

Le contexte qui amène l'étudiant à poser le problème statistique et la nécessité d'une interprétation doivent le conduire à introduire une intelligibilité au cours des traitements et à l'issue des traitements qui ne serait pas nécessaire d'un point de vue strictement mathématique.

Gestion de l'hôpital :

PROBLEME DE GESTION

Etudier la durée d'hospitalisation.

PROBLEME STATISTIQUE

Distribution bi-variée
Durée d'hospitalisation (en jours) et Activité (en activité / Sans activité professionnelle).

Question : y-a-t-il une relation entre les deux variables ?

Schéma 2 : Problème-type de statistiques descriptives à 2 variables.

La maquette s'adresse à des élèves n'ayant reçu qu'un enseignement de statistiques descriptives à 1 variable.

Elle a deux objectifs :

OBJECTIF 1 : APPRENDRE A ANALYSER UNE DISTRIBUTION STATISTIQUE ;

OBJECTIF 2 : CONSTRUIRE, A PARTIR DES RESUMES STATISTIQUES DEJA ETUDIES (ESSENTIELLEMENT LA MOYENNE ET LA VARIANCE) LES OUTILS STATISTIQUES SERVANT A CARACTERISER LA RELATION ENTRE DEUX VARIABLES.

CARACTERISTIQUES DE LA TACHE :

LE BUT :

Il s'agit de "dire quelque chose" sur la distribution, c'est-à-dire de parvenir à un **énoncé valide** ;

LE TRAITEMENT :

L'étudiant n'a pas encore les moyens de traiter la tâche proposée. Il lui faut néanmoins **utiliser les outils connus et se rendre compte de la limite de ses conclusions à partir des traitements qu'il a pu effectuer.**

CARACTERISTIQUES DU SUJET :

Lorsqu'il aborde ce type de tâches, l'étudiant a souvent du mal à retenir et à utiliser les liens qui existent entre les différents résumés statistiques (ex : la moyenne et la variance), tant leur calcul est encore coûteux. Il pense, d'autre part, que l'outil sera nouveau et non pas construit à partir de ceux étudiés lors de l'enseignement des statistiques à une variable.

III - CHOIX DE TROIS CLASSES DE PROBLEMES HIERARCHISEES

Tout problème fait intervenir deux variables pour lesquelles nous avons un ensemble de valeurs.

Nous avons retenu la distinction courante en statistiques appliquées aux Sciences Humaines entre **trois niveaux de mesures**. Chaque niveau est défini par les propriétés de l'ensemble des valeurs que l'on peut obtenir par des opérations de mesure.

Ces niveaux sont hiérarchisés, c'est-à-dire que les valeurs jouissent de toutes les propriétés du niveau inférieur plus d'autres propriétés.

NIVEAU 1 : ECHELLE NOMINALE

Création de classes d'équivalence

Il s'agit seulement de répartir les observations dans un ensemble de classes.

Exemples : catégories socio-professionnelles de l'I.N.S.E.E.

sexe (1 : garçons, 2 : filles).

NIVEAU 2 : ECHELLE ORDINALE

Etablissement d'une relation d'ordre entre les observations

Les observations peuvent être, à présent, ordonnées.

Exemple : réponses à une enquête

"Aimez-vous ce produit :

beaucoup ?

assez ?

moyennement ?

peu ?

pas du tout ?

NIVEAU 3 : ECHELLE D'INTERVALLES

La différence entre les différentes classes de mesures a un sens.

Exemples : durée d'hospitalisation (en jours) ;

âge (en années).

Nous ajoutons, à ce niveau, une distinction entre variables continues et variables discrètes.

Le croisement de ces trois types d'échelles pour les deux variables nous a permis de définir 6 classes de problèmes (cf tableau 1).

var 1 \ var 2	NOMINALE	ORDINALE	INTERVALLES
NOMINALE	Khi 2	Khi 2	Rapport de corrélation Analyse de variance
ORDINALE	/ / / / /	Coefficient de corrélation par rangs Coefficient de Spearman	Rapport de corrélation Analyse de variance Etude de la forme de la relation
INTERVALLES	/ / / / /	/ / / / /	Coefficient de corrélation Analyse de variance Courbe de régression

Tableau 1 : Les six classes de problèmes distinguées pour les problèmes de relation entre deux variables et les outils statistiques et graphiques correspondants.

Trois classes de problèmes ont été choisies : celles qui font intervenir au moins une variable d'intervalles.

Il s'agit donc des cas :

NOMINALE / INTERVALLES,

ORDINALE / INTERVALLES,

INTERVALLES / INTERVALLES.

Trois aspects nous semblent intéressants :

1. Nous désirons proposer aux étudiants des problèmes faisant intervenir des variables nominales.

En effet, il existe **déjà** pour les problèmes appartenant à la classe nominale / intervalles une solution intéressante.

Par ailleurs, manipuler des variables nominales s'avère plus difficile pour les étudiants :

- Nous constatons souvent la suprématie des outils quantitatifs, d'où une première difficulté à se situer à d'autres niveaux d'activité ;
- Les représentations graphiques ne peuvent **pas encore** se substituer à l'ensemble de la distribution statistique, mais tout au plus aux différentes classes (histogramme).

2. De nouveaux modes de représentation interviennent pour les échelles ordinale et d'intervalles et de nouvelles caractéristiques de la distribution prennent un sens.

Modes de représentation :

Pour les échelles d'intervalles, une représentation sous la forme d'un nuage de points peut se substituer à la distribution présentée comme un ensemble de mesures ;

Caractéristiques de la distribution :

Dès le niveau ordinal, la forme de la distribution a un sens.

Le problème de la relation entre les deux variables est donc enrichi par l'étude de la forme de celle-ci (niveau ordinal) ou de la recherche d'une courbe de regression (niveau intervalles) qui va parfois permettre de résumer graphiquement la distribution bi-variée.

3. Les outils statistiques construits pour la classe NOMINALE / INTERVALLES sont utilisables pour les autres classes de la hiérarchie. D'autre part, le rapport de corrélation fondera la construction du coefficient de corrélation.

III - ARCHITECTURE DU SYSTEME

L'architecture du système peut être représentée de la façon suivante :

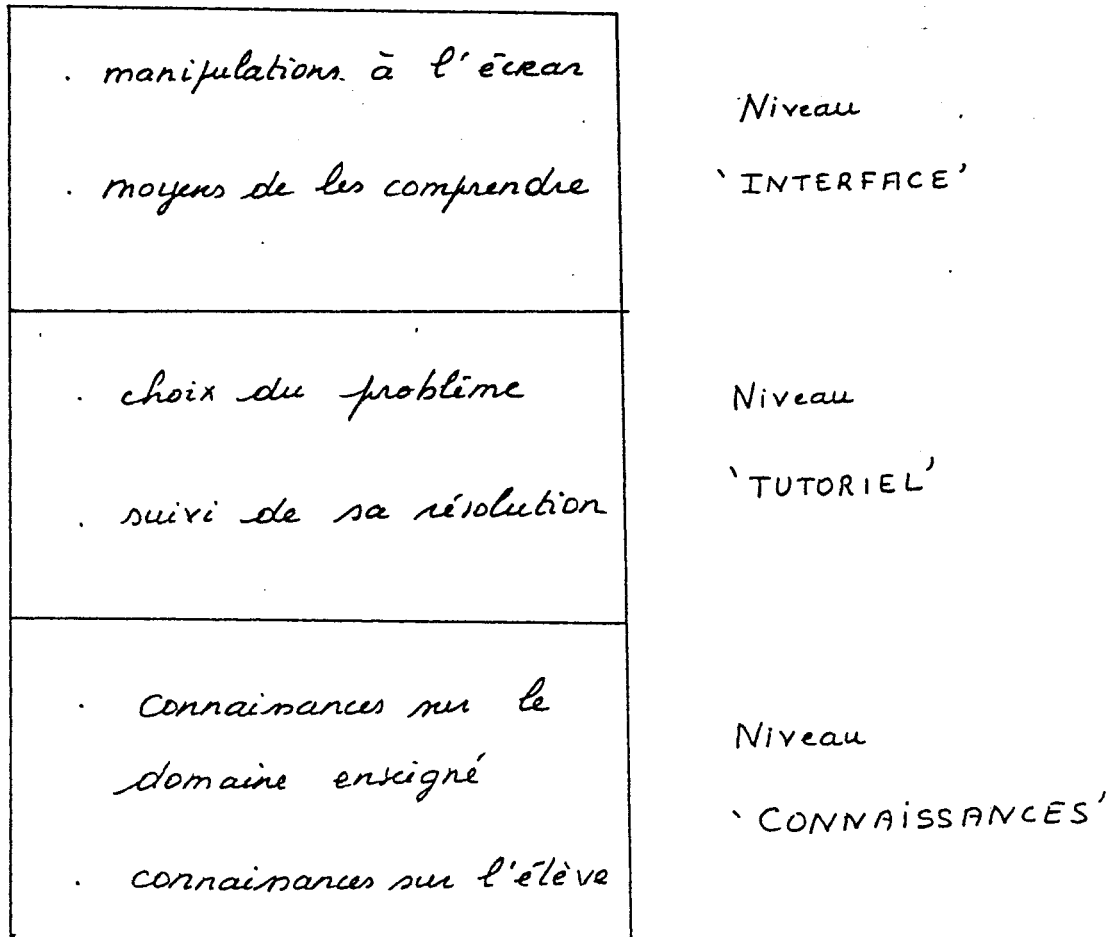


Schéma 3 : Structure générale du didacticiel.

Nous présentons le didacticiel selon trois niveaux :

3.1 - Le niveau "Interface"

Le travail de conception correspondant consiste à définir d'une part, les manipulations possibles à l'écran par l'élève et par la système et, d'autre part, les moyens de les comprendre.

3.2 - Le niveau "tutoriel"

Ce niveau comporte les fonctionnalités suivantes :

a) Le choix du problème à résoudre, selon :

- les connaissances sur le domaine enseigné que nous avons découpé en "classes de problèmes" ;
- les connaissances sur l'élève.

b) Le suivi de la résolution du problème

Il s'agit ici de concevoir un outil permettant de comprendre les actions de l'élève.

Nous entendons par compréhension des actions de l'élève, les deux aspects suivants :

- Inférer la démarche d'ensemble de l'élève si une démarche d'ensemble structure son activité de résolution ;
- Organiser ses actions sous la forme d'un plan.

Cette fonction nécessite également un diagnostic des difficultés de l'élève qui peuvent être de deux ordres :

- l'élève ne dispose pas des connaissances nécessaires,
- l'élève a du mal à organiser ses actions.

3.3 - Le niveau "Connaissances".

Deux types de connaissances sont distinguées :

- les connaissances sur le domaine des statistiques descriptives à deux variables.

Cet aspect correspond à un recueil d'expertise auprès d'un enseignant de statistiques.

- les connaissances sur les élèves.

Cet aspect exige la construction, en amont, d'un modèle qui distingue des "prototypes" d'élèves et décompose la résolution du problème en une série de "pas".

IV - REALISATION DU DIDACTICIEL : ETAT D'AVANCEMENT

Les trois niveaux de l'architecture du didacticiel définis en III interfèrent évidemment.

Notons cependant la place centrale du niveau "Tutoriel" qui est la couche dynamique du didacticiel (certains auteurs l'appellent d'ailleurs "moteur d'enseignement") et fait donc intervenir les deux autres niveaux.

Aussi l'étude réalisée sur la représentation des connaissances du domaine (niveau "connaissances", cf 3.3) que nous présentons dans la suite de ce texte intègre-t-elle, essentiellement par la distinction entre "démarches" et "actions" la caractéristique tutorielle de PROSAIC : la construction d'un outil capable de reconnaître une démarche ou d'aider à son élaboration, tout en s'adaptant à l'hétérogénéité initiale des utilisateurs potentiels.

C'est la conception de cet outil que nous abordons actuellement. Elle exige l'élaboration d'un modèle à deux composantes :

- la composante "Résolution d'un problème de relation entre deux variables statistiques par l'élève" (1) ;
- la composante "Compréhension des actions des élèves par les enseignants de statistiques" (2).

Ces deux composantes nécessitent la définition d'une stratégie de recherche faisant intervenir des méthodologies aussi différentes que :

- la résolution "collective" (mode "papier-crayon") de problèmes afin de réaliser une première analyse de protocoles et l'élaboration d'une première version de la composante "Résolution" (1).

Cette version devient alors le support de travail de l'équipe d'enseignants de statistiques qui doivent élaborer une première version de la composante "Compréhension des actions des élèves" (2).

- une série d'entretiens individuels à partir desquels la composante "Résolution" (1) est affinée et la composante "Compréhension" (2) validée lors de l'interaction entre le sujet et l'expérimentateur.

V - ETUDE REALISEE SUR LA REPRESENTATION DES CONNAISSANCES DU DOMAINE.

5.1 - Connaissances associées à un problème particulier.

Nous choisissons, dans cette partie du texte, de présenter les connaissances associées à la résolution d'un problème particulier.

Ce problème est le suivant :

Un directeur d'hôpital désire étudier la durée de séjour dans son établissement.

Il dispose de l'ensemble des observations (activité du patient, durée de séjour).

Existe-t-il une relation entre la variable activité et la variable durée de séjour ?

5.2 - Distinction entre trois résolveurs.

Nous distinguons trois résolveurs :

a - L'EXPERT EN STATISTIQUES, que nous appelons l'EXPERT A.

Il identifie les conditions de calcul du rapport de corrélation et le calcule.

b - L'ENSEIGNANT DE STATISTIQUES que nous appelons l'EXPERT B.

Il construit le rapport de corrélation avec ses élèves.

c - L'ELEVE

Il doit :

construire le rapport de corrélation (nouvel outil) à partir des paramètres connus antérieurement (la moyenne et la variance).

5.3 - Distinction entre trois types de connaissances : les objets, les actions et les démarches.

5.3.1 Choix d'une représentation orientée objet.

La conception de la partie "Connaissances du domaine" est facilitée par l'adoption d'une philosophie particulière : celle de la programmation orientée objet.

Quatre principes caractérisent cette philosophie :

- a. Il n'y a qu'une sorte d'entité : l'objet, composé :
 - de variables qui caractérisent l'état de l'objet ;
 - de méthodes qui permettent d'utiliser les caractéristiques de l'objet.

- b. Tout objet est une instance d'une classe.
Ainsi, toutes les distributions nominales vont être regroupées au sein de la classe NOMINALE.

- c. Tout objet est activé à la réception d'un message.
exemple : distribution calculer-mode.

- d. Toute classe est sous-classe d'une autre classe et la sous-classe hérite des propriétés des ancêtres.
exemple : la classe ORDINALE
est une sous-classe de la classe NOMINALE

Dans la conception de la partie "Connaissances sur le domaine", les objets scientifiques, les actions et les démarches seront tous représentés sous la forme d'objets selon les quatre principes donnés ci-dessus.

5.2 - La structure générale de la partie "Connaissances sur le domaine".

5.2.1 Utilisateur et environnement statistique

Nous distinguons deux composantes :

- des connaissances portant sur un environnement statistiques caractérisé par des objets statistiques, des actions et des démarches ;
- des problèmes posés par l'utilisateur (qu'il soit humain ou non).

Le lien entre les deux éléments de ce couple (environnement, utilisateur) comporte deux aspects :

- a) les objets effectivement utilisés qui sont des réalisations particulières de la classe à laquelle ils appartiennent.

exemple :

Pour les objets statistiques, la distribution bivariée Activité et Durée de séjour est une instance de Distribution bivariée.

Les actions et les démarches effectivements utilisées donnent lieu, elles aussi, à la création d'instances.

b) la construction d'un "réseau procédural" c'est-à-dire de la séquence d'actions utilisée pour résoudre le problème :

- ces actions sont en fait des instances d'actions (actions effectivement utilisées) ;
- ces actions peuvent être détaillées : le réseau peut correspondre à une série de plans hiérarchisés.

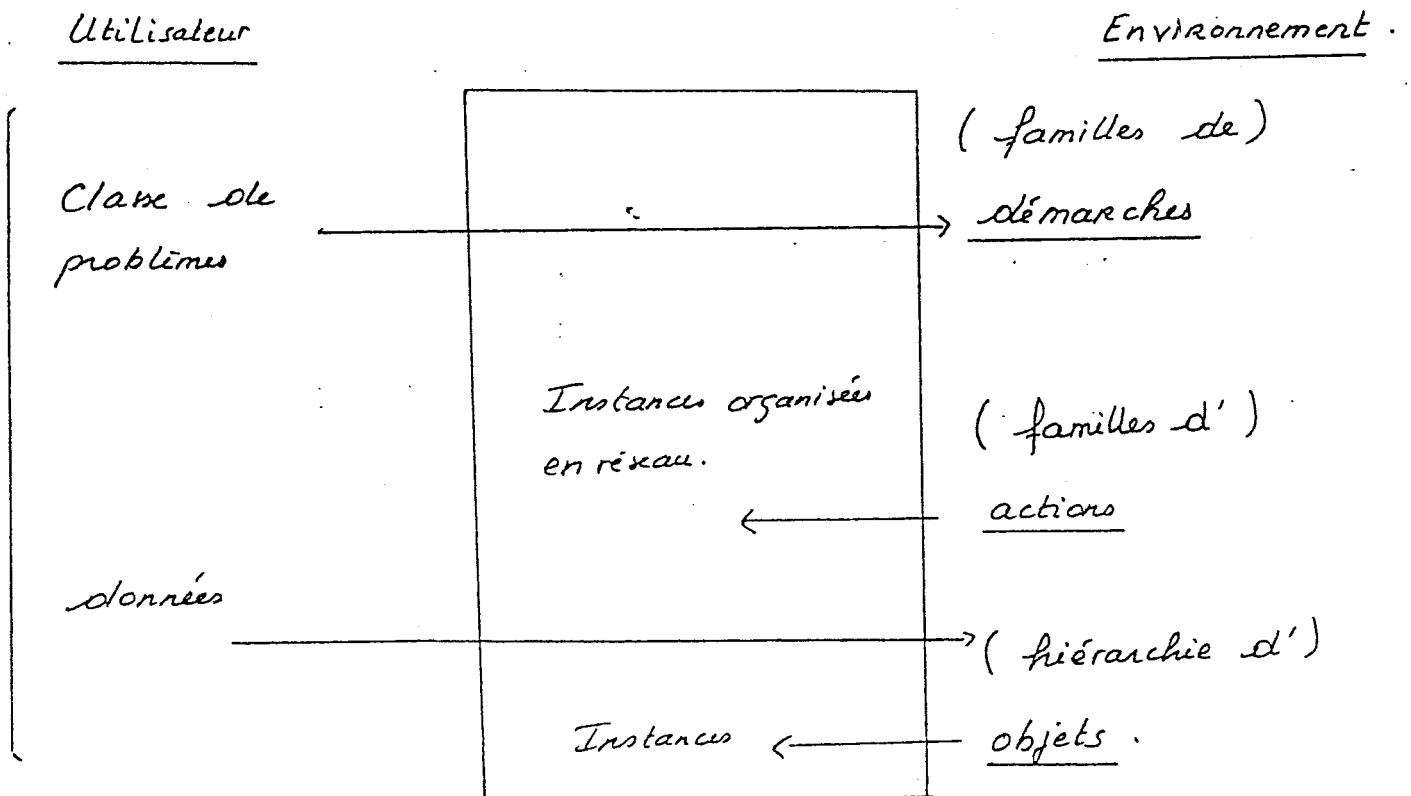


Schéma 4 Liaison Utilisateur - Environnement.

5.2.2 - Résolution du problème par l'expert A (l'expert en statistiques).

Nous allons caractériser la résolution du problème "activité et durée de séjour" par l'EXPERT A et sa correspondance informatique.

a) Création des instances correspondant à la distribution bi-variée "Activité et durée de séjour" et aux distributions uni-variées "Activité et durée de séjour".

Les "objets" sont les équivalents informatiques des objets statistiques que sont les distributions, les données, les représentations graphiques et les tableaux.

Nous avons, pour l'instant, proposé la hiérarchie suivante :

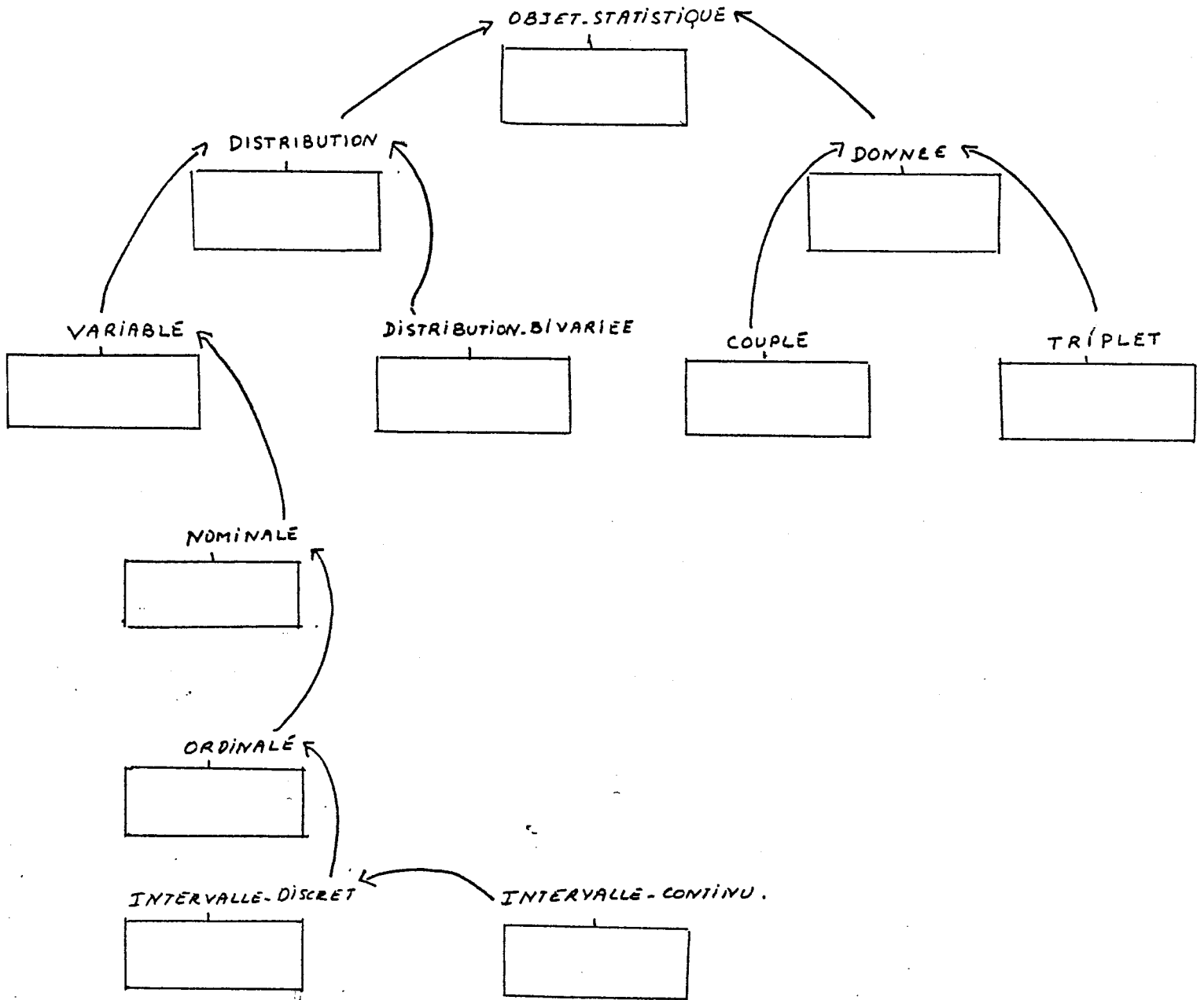


Schéma 5 : Hiérarchie d'objets statistiques.

- La distribution bi-variée "Activité et durée de séjour" est créée comme instance de DISTRIBUTION-BIVARIEE,
- Les distributions "Activité" et "Durée de séjour" sont toutes deux créées comme instances de VARIABLE car leur nature n'a pas encore été identifiée,

b) Résolution du problème

Les actions et les démarches ont la même structure, à une caractéristique près :

- une démarche ne peut être exécutée mais seulement détaillée en une suite d'actions ;
- une action peut être exécutée ou détaillée en une suite d'actions.

Les actions et les démarches sont regroupées par familles.

Ainsi, pour la résolution du problème de relation entre une variable NOMINALE et une variable d'INTERVALLES allons-nous avoir besoin de :

- Trois familles d'Actions.

1. La famille NATURE-VARIABLES qui permet de reconnaître la nature d'une variable statistique (Nominale, Ordinale, Intervalle-discret ou Intervalle-continu) ;

2. La famille CARACTERISER qui correspond aux actions de calcul des paramètres statistiques ;

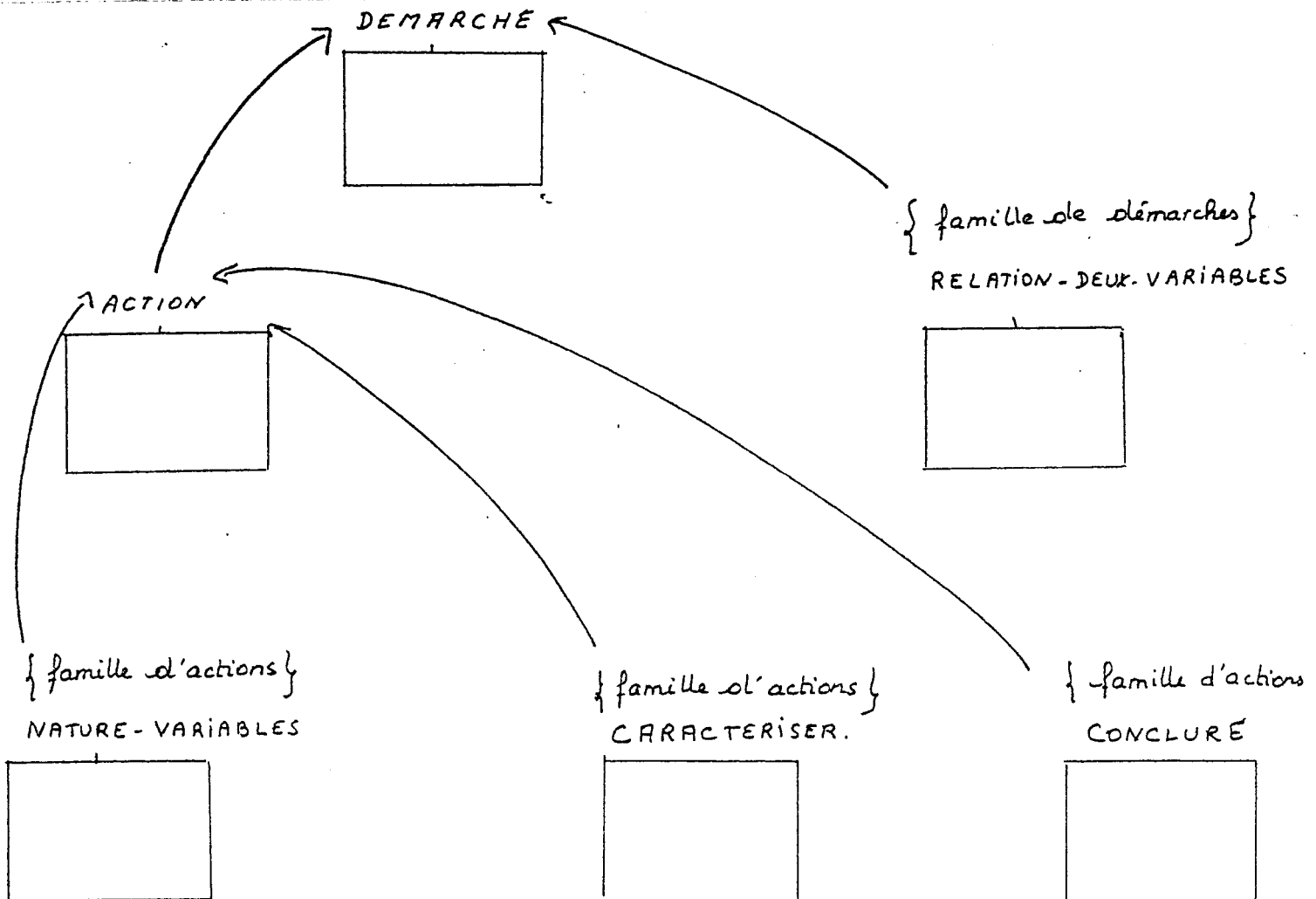
3. La famille CONCLURE qui correspond aux actions de conclusion.

- Une famille de démarches.

Il s'agit de la famille RELATION-DEUX-VARIABLES.

Elle permet la résolution du problème de relation entre deux variables pour les six classes de problèmes.

Nous proposons donc la structure générale suivante :



Comment cette structure est-elle utilisée ?

La démarche RELATION-DEUX-VARIABLES-DIS, qui appartient à la famille RELATION-DEUX-VARIABLES reçoit le message "Méthode de construction du plan, (X. Y) (X et Y sont à associer aux variables activité et durée de séjour).

La méthode de construction du plan n'est pas directement reconnue par RELATION-DEUX-VARIABLES-DIS, mais elle est recherchée dans la hiérarchie et finalement retrouvée dans le prototype-démarche DEMARCHE.

Nous avons défini le réseau de résolution comme des instances d'actions ou de démarches.

Comment la méthode de construction du plan va-t-elle "travailler" ici ?

a) Le noeud correspondant au but du problème proposé est créé.

Ce noeud est une instance de RELATION-DEUX-VARIABLES-DIS pour les variables Activité et Durée de séjour.

- Le champ variable : il faut avoir une variable "au moins" nominale et une variable "au moins" intervalle-discret.

La famille d'actions NATURE-VARIABLES interviendra car Activité et Durée de séjour sont encore des instances de VARIABLE

A l'issue de cette vérification, Activité sera une instance de NOMINALE et Durée de séjour une instance d'INTERVALLE-CONTINU.

- Le champ contexte : désire -t-on effectivement étudier la relation entre les deux variables sous l'angle de la dispersion.

Un dialogue avec l'utilisateur intervient ici.

- Le champ pré-condition : il porte ici sur la variable principale (Durée de séjour) et l'exigence d'une dispersion des données importantes.

Cette exigence est ici traduite par un test sur la valeur du coefficient de variation.

b) Une fois ces vérifications effectuées, le champ conséquent indique la construction du plan de résolution à partir d'un autre champ le champ body.

Il y aura alors création de deux noeuds :

- le premier noeud est une instance de l'action RAPPORT-DE-CORRELATION qui appartient à la famille CARACTERISER ;

- le second noeud est une instance de l'action CONCLURE-SEUIL qui appartient à la famille CONCLURE.

5.2.3 - Objets, actions et démarches : une analyse de l'environnement statistique.

Le travail sur la représentation des connaissances repose sur une analyse de l'environnement statistique.

a) Les objets

La hiérarchie informatique proposée reproduit la hiérarchie existant dans le domaine statistique entre les échelles et les types de variables associés.

Lorsque nous aborderons des classes de problèmes pour lesquelles la forme de la distribution intervient, un problème intéressant se posera : le partage des mêmes caractéristiques par des objets statistiques et graphiques.

b) Les démarches

La famille "RELATION-DEUX-VARIABLES" correspond à l'analyse que nous avons réalisée :

- sur les classes de problèmes (rappelons que par le croisement sur la nature des variables, nous en avons distingué six) ;
- sur les caractéristiques de la relation qui peuvent être étudiées (les fréquences, la dispersion, la forme).

c) Les actions

C'est la partie des connaissances qui nous intéresse le plus actuellement.

L'idée de "familles" d'actions repose sur la distinction relative entre buts et actions. (les actions au sens propre étant, elles, exécutables) et sur l'importance des contextes et de leur analyse.

Ainsi, CONCLURE constituera une famille d'actions qui aura la structure suivante en statistiques descriptives :

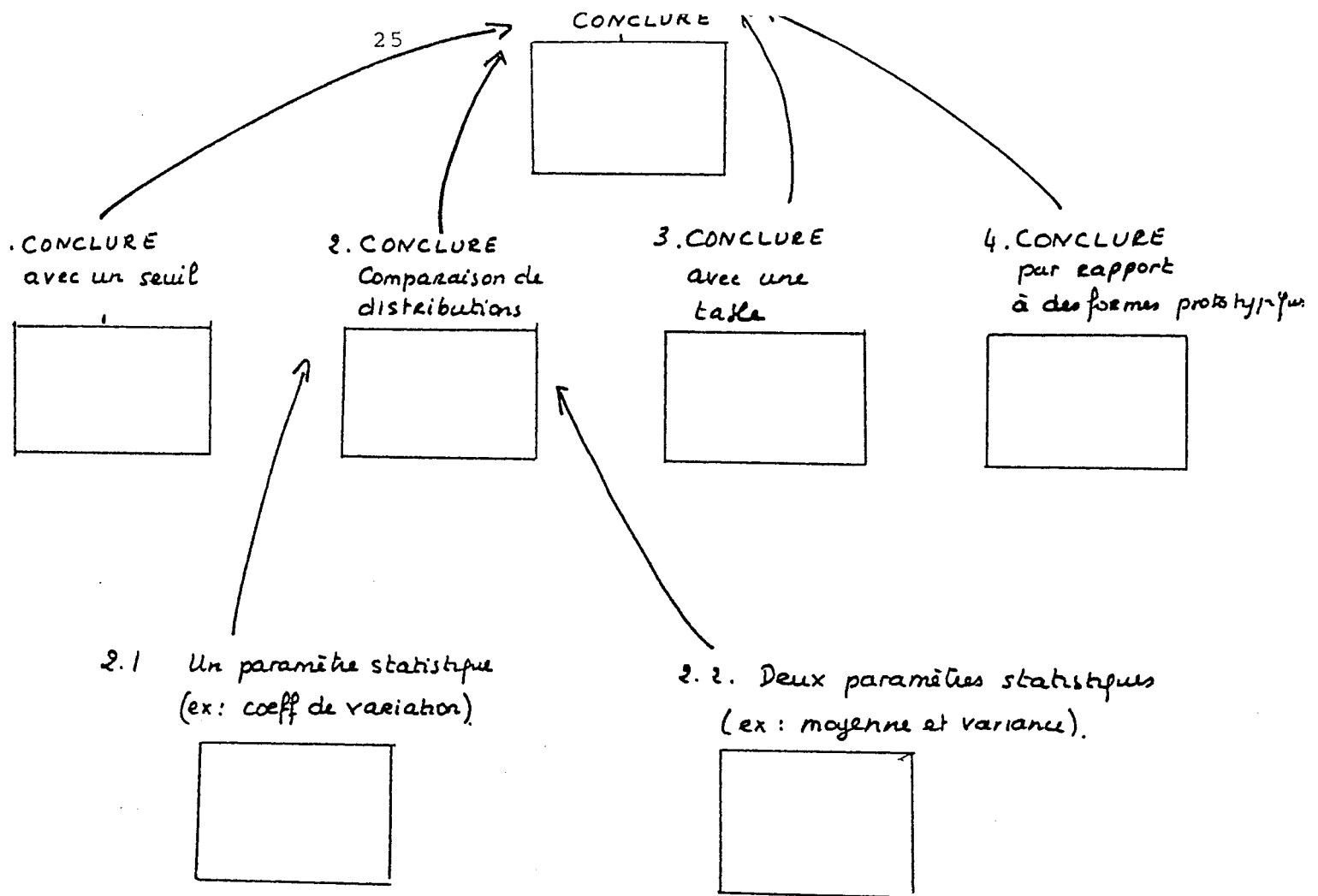


Schéma 7 : L'exemple de la famille d'actions 'CONCLURE'

Nous pensons que c'est par la définition de familles d'actions que sont introduits, dans la conception de PROSAIC, l'enseignant de statistiques (EXPERT B) et l'élève :

- La démarche proposée par l'EXPERT B fait intervenir non seulement des actions (au sens "exécutables") mais aussi des buts plus généraux.

Ainsi, le problème de la relation entre les deux variables "activité" et "durée de séjour" peut-il être reformulé de la façon suivante : "il faut diminuer la dispersion observée". Construire le rapport de corrélation est un moyen d'atteindre ce but, une fois le contexte bien analysé.

- Aussi, un travail important consiste à définir les grandes familles d'actions dont l'expert A, l'expert B et l'élève peuvent avoir besoin (celles qui concernent l'expert A étant moins nombreuses).

Il s'agit actuellement des six familles suivantes :

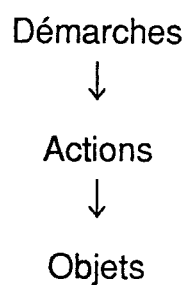
1. SUBSTITUER
2. CARACTERISER
3. COMPARER
4. ORGANISER
5. CONCLURE
6. NATURE-VARIABLES

Nous donnons ici les noms des six familles qui sont aussi les racines d'autant de sous-arborescences de l'arborescence dont le prototype ACTION est la racine.

VI - CONCLUSION

La prochaine étape de la conception de PROSAIC est l'élaboration de l'outil de compréhension des actions de l'élève en situation de résolution de problèmes.

Nous avons présenté dans la 5ème partie de ce texte la structure "Verticale" des connaissances :



Le processus de compréhension est en fait inverse puisqu'il part des actions de l'élève et non des démarches.

Les auteurs (essentiellement américains) d'articles sur les méthodes de planification et de reconnaissance de plans opposent sur ce point les deux outils : l'outil de planification ("Planner") et l'outil de reconnaissance ("Plan Recognizer").

L'outil de compréhension que nous devons concevoir pour PROSAIC est cependant différent d'un "plan recognizer".

Il s'agit :

- de tâches "mal définies" puisque c'est la conception d'outils statistiques nouveaux qui est en jeu ;

- de tâches pour lesquelles nous pensons que l'élève se définira en séquence, des buts à atteindre, la nature de ces buts étant différente de celle de l'EXPERT A (nous pensons essentiellement ici aux familles SUBSTITUER et ORGANISER);