

UNIVERSITÉ DE NEUCHÂTEL
FACULTÉ DE DROIT ET DES SCIENCES ÉCONOMIQUES

Conception et étude de la méthodologie de développement d'un système de Formation Assistée par Ordinateur

Application dans le domaine de la statistique

THÈSE

présentée à la Faculté de Droit et des Sciences économiques
de l'Université de Neuchâtel pour l'obtention du grade de
docteur ès sciences économiques

par

Nicole Rebetez

Imprimerie de l'Évole, Neuchâtel
1993

Mademoiselle Nicole REBETEZ est autorisée à imprimer sa thèse de doctorat ès sciences économiques, intitulée: "Conception et étude de la méthodologie de développement d'un système de Formation Assistée par Ordinateur".

Elle assume seule la responsabilité des opinions énoncées.

Neuchâtel, le 15 février 1993

Le doyen
de la Faculté de droit
et des sciences économiques

H.-R. Schüpbach

Remerciements

Je remercie sincèrement mon directeur de thèse le Professeur Yadolah Dodge pour ses conseils, ses encouragements et son soutien pendant toute la durée du travail.

Mes remerciements vont également au Professeur François Hainard, co-rapporteur de cette thèse, ainsi qu'au Professeur Jacques Savoy et à M. Thierry Gafner pour leur lecture critique de ce travail.

Je remercie le Dr.P.A.Jaeger et M. Hansjürg Neuhaus de la Fondation ASCOM-Hasler pour l'intérêt qu'ils ont porté à la réalisation de ce projet et l'aide constructive qu'ils m'ont apportée.

Je remercie chaleureusement tous mes collègues du Groupe de Statistique ainsi que les collaborateurs du Groupe d'Informatique qui ont participé de près ou de loin à la réalisation de ce travail, pour leurs conseils, leur aide et leur disponibilité. Mes remerciements vont tout particulièrement à Mme Sylvie Gonano-Weber, qui fut ma collègue durant toute cette période et qui a participé activement à l'aboutissement de cette thèse, par son travail assidu, ses conseils précieux, et son soutien moral indispensable.

Mes remerciements vont encore à toute ma famille et spécialement à ma maman pour son aide dans la relecture de la thèse, ainsi qu'à Pierre-Alain pour ses encouragements précieux.

Je remercie finalement le Fond National Suisse pour la Recherche Scientifique (FNRS) et la Fondation ASCOM-Hasler pour leur soutien financier.

A mes parents

Préface

Les méthodes statistiques sont utilisées dans une très large variété de contextes, aussi divers que la recherche médicale, recherche et expériences sur les causes, le diagnostic, le traitement et la prévention des maladies; le contrôle de qualité; les recherches en sciences économiques, politiques ou sociales; les études démographiques, croissance des populations, mouvements migratoires; le domaine des assurances, détermination des taux d'accidents, des taux de primes; le marketing, études de marché, études de préférences des consommateurs et de leurs habitudes; la santé, problèmes des soins médicaux, de l'hospitalisation, des assurances et de la santé publique, etc. Le champ d'application des méthodes statistiques est tellement vaste qu'il serait impossible d'en fournir une liste exhaustive!

Pour faire face aux besoins, de nombreuses écoles et universités dispensent des cours en vue de familiariser les étudiants avec les techniques statistiques les plus utilisées. Toutefois, nous pouvons remarquer que de nombreux utilisateurs d'outils statistiques n'ont pas les connaissances et les compétences nécessaires à la résolution des problèmes auxquels ils sont confrontés.

Les logiciels statistiques offrent généralement un ensemble d'outils que l'utilisateur ne fait qu'appliquer sans connaître les fondements théoriques de ces méthodes, les hypothèses sous-jacentes à leur utilisation, leur cadre d'application ni même la manière correcte d'interpréter les résultats fournis.

L'utilisation des techniques statistiques requiert non seulement des connaissances théoriques, mais également la compréhension globale des notions, de leur utilisation et de leurs limitations. Une des voies d'acquisition de cette compréhension est l'expérience, c'est-à-dire l'essai, le test, le jeu.

Notre objectif est d'étudier la possibilité d'offrir une méthode qui tente de lier l'apprentissage théorique des notions avec la possibilité pour l'utilisateur de tester les notions apprises avec ses propres exemples, ses propres données. Cette méthode doit permettre à l'utilisateur, ou l'étudiant, d'acquérir des connaissances de manière active et attractive.

La méthode que nous proposons doit permettre au néophyte d'approcher le domaine soit au travers d'un **dictionnaire électronique** lui permettant d'accéder à des informations spécifiques telles que la définition, l'historique, les aspects mathématiques relatifs à un mot particulier, soit à l'aide d'un **livre électronique** lui permettant de découvrir pas à pas les différents domaines de la statistique.

De plus, nous proposons de lier ces deux approches de l'apprentissage "théorique" à une approche beaucoup plus pratique, consistant en un ensemble d'**outils statistiques** mis à la disposition de l'utilisateur et lui permettant de mettre ses connaissances à l'épreuve en testant les méthodes exposées par le texte avec ses propres données.

On assiste aujourd'hui à une généralisation de l'utilisation de l'outil informatique à tous les métiers. De nombreuses recherches ont notamment été réalisées ces dernières années dans son application au domaine de l'enseignement. L'informatique en tant qu'outil d'aide à l'enseignement a donné naissance à la nouvelle technologie de l'**Enseignement Assisté par Ordinateur**. Toutefois, comme le soulignait Ch. Duchateau en 1986 [14], "la situation me paraît stagnante: on en parle énormément, on en voit fort peu". Aujourd'hui, les développements se font de plus en plus nombreux, mais le domaine de la statistique n'est encore que très peu touché par cette technologie. Nous avons voulu contribuer à combler ce manque par le développement d'un système de Formation Assistée par Ordinateur dans ce domaine particulier.

Le but de cette thèse est de présenter les fruits de nos recherches concernant la méthodologie de construction de ce système. L'ensemble des considérations techniques ou générales que nous présentons est lié au travers de l'ouvrage avec l'application pratique relative à l'enseignement dans le domaine de la statistique que nous avons développée.

Nous présentons dans le chapitre 1 quelques éléments de base à la compréhension de l'ensemble du texte. La première section est un bref exposé sur les notions d'apprentissage et d'enseignement. Dans la section 2, nous proposons quelques considérations sur l'utilisation de l'ordinateur en situation de formation: comment et pourquoi l'ordinateur a-t-il pris son essor dans ce domaine? La section 3 décrit ce qu'il faut comprendre par la notion de "système de Formation Assistée par Ordinateur". Dans la section 4, nous attirons l'attention sur les différents types de connaissance à la base du développement d'un tel système. La section 5 en décrit succinctement les différentes étapes de développement. La section 6 enfin précise pourquoi notre développement

s'est axé sur le domaine de la statistique: pourquoi ce domaine est-il particulièrement bien adapté à un enseignement informatisé?

Le corps de la thèse est composé de deux parties. La première partie présente les différentes étapes de la **conception** d'un outil de Formation Assistée par Ordinateur au niveau **pédagogique**. Elle vise à mettre en évidence les besoins des utilisateurs auxquels celle-ci est destinée à répondre, à les exprimer de manière suffisamment précise pour aboutir à la description d'un produit conforme à ces besoins.

La première étape, appelée **analyse pédagogique**, est développée dans le chapitre 2. Elle a pour but de poser clairement les bases du problème afin de déterminer de manière détaillée **où l'on va**. C'est dans cette phase que nous établirons une description précise de nos **objectifs**, sur la base de l'étude des références disponibles; du public potentiellement touché par l'application; des types de besoin auxquels l'application est destinée à répondre; des types d'apprentissage qui en découlent; et également de la nature et de l'étendue des connaissances à transmettre. Cette étape d'analyse pédagogique permettra d'aboutir à une première description globale de l'application. Les objectifs principaux ayant été déterminés, nous serons en mesure d'explicitier la **stratégie pédagogique** soutenant l'application.

Le **scénario pédagogique** présenté au chapitre 3 permettra ensuite de définir les **moyens** d'atteindre les objectifs fixés, tant au niveau des moyens informatiques à mettre en oeuvre que du potentiel scientifique relatif au domaine traité.

La dernière étape de la conception pédagogique est d'aboutir à une **description de l'application** sur laquelle nous pourrions nous baser en vue de concevoir son adaptation informatique. Celle-ci est présentée au chapitre 4.

La seconde partie de la thèse développe les concepts **informatiques** de base au développement de l'application. Le but de cette partie est d'accéder à un niveau plus pratique, celui du développement informatique des programmes permettant de construire et de gérer l'application telle que nous l'avons conçue.

Le chapitre 5 est une introduction qui met en évidence les différents environnements dans lesquels s'inscrit notre application. Les chapitres 6, 7 et 8 présentent une analyse informatique séparée du dictionnaire électronique, du livre électronique et des outils statistiques.

Liste des figures

1.1	Environnement d'apprentissage tuteur-étudiant	2
1.2	Types de connaissances nécessaires à la construction d'un système de Formation Assistée par Ordinateur (FAO)	12
1.3	Processus de production d'un logiciel (Source: Sommerville I. (1985). <i>Software Engineering</i> , second edition. Addison-Wesley Publishing Company)	14
2.1	Représentation schématique de la base de connaissances	29
2.2	Structure globale de l'application statistique	34
2.3	Positionnement du système dans l'espace des stratégies possibles	40
3.1	Interface homme-machine	43
3.2	Chaînage linéaire	48
3.3	Chaînage individualisé	49
3.4	Chaînage modulaire	50
3.5	Séquence des opérations relatives à un dictionnaire	51
3.6	Séquence des opérations relatives à un livre	53
3.7	Ecran à zones de commandes superposées (Source: <i>Prise en main de Framework III</i> , La Commande Electronique, 1988)	55
3.8	Ecran à zone de commandes fixe (Source: <i>Microsoft Word, guide d'utilisation</i> , Microsoft Corporation, 1989)	56
3.9	Ecran à fenêtres fixes (Source: Ronchetti E., Antille G. et Polla M. <i>STEPI, STatistique et Probabilités: une Introduction</i> , Presses Académiques, Neuchâtel, 1989)	62
3.10	Ecran à fenêtres superposées (Source: <i>Prise en main de Framework III</i> , La Commande Electronique, 1988)	62
3.11	Structure séquentielle	68
3.12	Structure hiérarchique	68
3.13	Structure en réseau	69
4.1	Division de l'écran en zones	79
4.2	Page-écran standard d'un outil statistique	84
4.3	Outils statistiques: écran texte complémentaire (1)	85
4.4	Outils statistiques: écran texte complémentaire (2)	85

4.5	Outils statistiques: écran graphique complémentaire	86
4.6	Représentation schématique du contenu du dictionnaire électronique	91
4.7	Page-écran du dictionnaire électronique	92
4.8	Index alphabétique	93
4.9	Index hiérarchique	94
4.10	Contexte du mot "loi de probabilité continue"	95
4.11	Seconde forme de la zone de commandes du dictionnaire électronique	95
4.12	Représentation schématique du livre	98
4.13	Représentation schématique d'une section	99
4.14	Table des matières du livre électronique (sommaire)	101
4.15	Table des matières du livre électronique (détail)	101
4.16	Index du livre électronique (sommaire)	102
4.17	Index du livre électronique (détail)	103
4.18	Page-écran du livre électronique	104
5.1	Environnement d'une application de consultation	111
5.2	Environnement d'une application interactive de calcul	114
6.1	Représentation matricielle de la base de données du dictionnaire électronique	118
6.2	Fichier de base du dictionnaire électronique	121
6.3	Accès aux fichiers DICOx par l'index	122
6.4	Processus de génération du dictionnaire électronique	124
6.5	Structure d'un fichier texte (input) du dictionnaire	126
6.6	Terminologie associée à une structure hiérarchique	129
6.7	Menus de l'interface opérateur du dictionnaire électronique	131
6.8	Processus d'interrogation du dictionnaire électronique	134
6.9	Temps d'accès par une recherche dichotomique	136
7.1	Numérotation effective des sections	147
7.2	Numérotation séquentielle des sections	147
7.3	Terminologie associée à une structure hiérarchique (2)	149
7.4	Processus de génération du livre électronique	152
7.5	Structure d'un fichier texte (input) du livre électronique	154
7.6	Représentation de l'index du livre électronique	156
7.7	Menus de l'interface opérateur du livre électronique	157
7.8	Processus de consultation du livre électronique	161
8.1	Schéma d'appel d'un outil statistique	169

8.2	Fenêtre de présentation des résultats	173
8.3	Masque de saisie pour des données groupées par valeur	175
8.4	Masque de saisie pour des données brutes relatives à deux échantillons	175
8.5	Parcours logique de saisie de données groupées par valeur	176
8.6	Parcours logique de saisie de données brutes relatives à deux échantillons	176

Table des matières

1	Introduction	1
1.1	Apprentissage et enseignement	1
1.2	L'outil informatique en situation de formation	5
1.2.1	Un peu d'histoire	5
1.2.2	Avantages apportés par l'ordinateur	6
1.2.3	Réticences vis-à-vis de l'ordinateur	8
1.3	Richesse d'un système de Formation Assistée par Ordinateur	9
1.4	Connaissances nécessaires à la construction d'un système de Formation Assistée par Ordinateur	11
1.5	Processus de production d'un système de Formation Assistée par Ordinateur	13
1.6	Spécificités du type d'application	16
I	CONCEPTION PEDAGOGIQUE	19
2	Analyse pédagogique	21
2.1	Introduction	21
2.2	Types de références disponibles	22
2.3	Public potentiel	24
2.4	Types de besoins	27
2.5	Types d'apprentissage	29
2.6	Détermination du contenu	30
2.6.1	Le vaste domaine de la statistique	30
2.6.2	Etendue et nature des sources d'information retenues	31
2.7	Description globale de l'application	32
2.8	Stratégie pédagogique	35
3	Scénario pédagogique	41
3.1	Introduction	41
3.2	Interface homme-machine	42
3.3	Séquence des opérations	46
3.3.1	Politiques de transmission de l'information	46

3.3.2	Chaînage linéaire	47
3.3.3	Chaînage individualisé	48
3.3.4	Chaînage modulaire	49
3.3.5	Séquence des opérations relatives à un dictionnaire électronique	50
3.3.6	Séquence des opérations relatives à un livre électronique	52
3.4	Dispositifs d'entrées	53
3.4.1	Choix du dispositif de commandes	53
3.4.2	Présentation des commandes	54
3.5	Dispositifs de sorties	56
3.5.1	Composition d'un écran	57
3.5.2	Utilisation des couleurs	58
3.5.3	Utilisation des graphiques	60
3.5.4	Utilisation du fenêtrage	60
3.6	Temps de réponse	61
3.7	Traitement des erreurs	64
3.8	Structuration du contenu	66
3.8.1	Relations internes	67
3.8.2	Construction des relations internes	70
3.8.3	Relations externes	74
4	Description de l'application	77
4.1	Introduction	77
4.2	Présentation de l'information	79
4.3	Outils statistiques	81
4.3.1	Objectifs des outils statistiques	81
4.3.2	Dispositifs de dialogue	82
4.3.3	Domaines couverts par les outils statistiques	86
4.4	Dictionnaire électronique	88
4.4.1	Objectifs du dictionnaire électronique	88
4.4.2	Contenu du dictionnaire électronique	89
4.4.3	Dispositifs de dialogue	90
4.5	Livre électronique	96
4.5.1	Objectifs du livre électronique	96
4.5.2	Contenu du livre électronique	97
4.5.3	Dispositifs de dialogue	100

II	ANALYSE INFORMATIQUE	107
5	Analyse informatique: architecture globale	109
5.1	Introduction	109
5.2	Application de consultation	110
5.3	Application interactive de calcul	113
5.4	Outils de développement	113
6	Analyse informatique du dictionnaire électronique	117
6.1	Introduction	117
6.2	Structure de la base de données	118
6.3	Description du programme de génération de la base de données	123
6.3.1	Génération des fichiers de base	125
6.3.2	Gestion des accès	127
6.3.3	Gestion de l'index hiérarchique	128
6.3.4	Interface opérateur	130
6.4	Méthodes d'accès au dictionnaire électronique	133
6.4.1	Recherche dichotomique	135
6.4.2	Recherche de sous-chaînes	137
6.4.3	Parcours d'arbre	139
7	Analyse informatique du livre électronique	143
7.1	Introduction	143
7.2	Structure de la base de données	144
7.2.1	Description des chapitres	144
7.2.2	Description des sections	145
7.2.3	Description des pages	148
7.2.4	Description du texte	150
7.3	Description du programme de génération de la base de données	150
7.3.1	Génération des fichiers de base du livre	151
7.3.2	Gestion de l'index du livre	153
7.3.3	Interface opérateur	157
7.4	Consultation du livre électronique	160
8	Analyse informatique des outils statistiques	167
8.1	Introduction	167
8.2	Accès aux outils statistiques	168
8.3	Structure d'une routine	170
8.3.1	Dispositifs de sortie des résultats	171

8.3.2 Dispositifs d'entrée de données 174

Conclusions **181**

Bibliographie **185**

Chapitre 1

Introduction

1.1 Apprentissage et enseignement

Nous considérons la notion d'apprentissage comme tout processus permettant à l'homme d'acquérir un ensemble de connaissances, quel que soit:

- le moteur de transmission de cette connaissance: un enseignant, ou un support matériel tel qu'un livre, une machine, un ordinateur;
- le mode d'acquisition de cette connaissance: un travail intellectuel ou l'expérience.

Commençons par placer la notion d'apprentissage dans son environnement. La figure 1.1 représente le cas d'un apprentissage avec un tuteur humain,

où les deux systèmes en présence sont d'une part le tuteur et d'autre part l'étudiant.

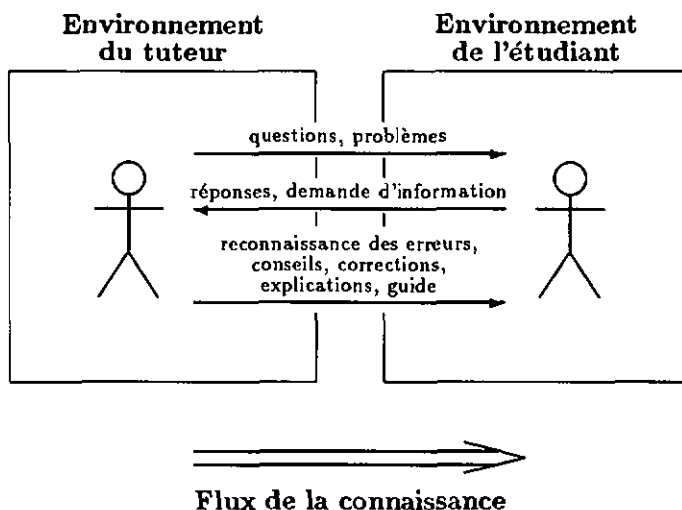


Figure 1.1: Environnement d'apprentissage tuteur-étudiant

La relation entre le tuteur et l'étudiant est une relation bilatérale de dialogue dans laquelle le tuteur joue le rôle de moteur de transmission de la connaissance. L'action principale de l'étudiant est l'assimilation de la connaissance, que ce soit par un processus de mémorisation, par la résolution de problèmes ou par le jeu.

L'étudiant possède généralement un support écrit (livre) pouvant contenir une discussion générale du domaine traité, des exemples lui permettant de mieux comprendre la matière et de tester sa compréhension, des problèmes ou des exercices à résoudre.

Ce support peut être relativement complet, auquel cas l'étudiant pourrait en principe le parcourir sans aide extérieure. Toutefois, il peut être confronté à des difficultés considérables:

- manque de volonté: l'étudiant abandonne car il n'est pas suivi, pas guidé;

- **passivité**: la lecture seule d'un livre demande une attention passive de la part de l'étudiant. Il préférera se dire qu'il a compris la matière plutôt que d'effectuer consciencieusement tous les exercices.

Le rôle du tuteur est alors de stimuler le travail de l'étudiant, de le guider dans son apprentissage, de reconnaître ses erreurs, de lui donner des conseils ou des explications lorsque cela est nécessaire. L'ensemble de ces tâches et la manière dont elles sont assumées forment un système d'**enseignement**.

Face à cet enseignement, le rôle de l'étudiant est un rôle d'**apprentissage**. Ces deux notions sont intimement liées, car interdépendantes: chaque méthode, ou chaque style d'enseignement permet une certaine forme d'apprentissage. Toute méthode d'enseignement doit donc être conçue en fonction du type d'apprentissage que l'on veut permettre.

Skinner [48] a mis en évidence trois théories classiques, qui mettent chacune l'accent sur un élément essentiel de l'apprentissage:

1. "nous n'apprenons que ce que nous faisons": cette théorie se base sur le fait qu'un élève n'absorbe pas passivement le savoir, mais doit jouer un rôle actif dans son apprentissage;
2. "nous apprenons par expérience": la première théorie est insuffisante, car il ne suffit pas toujours de faire quelque chose pour apprendre. Cette deuxième théorie met l'accent sur le monde, ou l'environnement dans lequel se présente le phénomène à apprendre, l'expérience étant comprise ici comme une condition, ou un milieu d'expérimentation. L'élève doit alors être capable de reconnaître ce milieu afin d'organiser ses actions;
3. "nous apprenons par essais et erreurs": cette troisième théorie apporte un nouvel élément: celui des conséquences liées au résultat de l'action. Une simple reconnaissance des erreurs permet d'éliminer les comportements erronés. Elle ne permet toutefois pas toujours de déterminer quel est le comportement correct.

Selon Skinner, ces trois théories ne peuvent être envisagées isolément, et il importe de les prendre toutes trois en considération dans toute situation d'apprentissage.

Nous ne tenterons pas d'effectuer une combinaison de ces théories afin de fournir une conception satisfaisante de l'apprentissage, car il serait vain de vouloir aboutir à une méthode d'apprentissage idéale et universelle.

Un des facteurs principaux permettant de concevoir une méthode d'apprentissage adaptée à une situation particulière est la matière enseignée, ou plus précisément le type de connaissances à transmettre. Voici quelques exemples qui mettent en évidence le type de connaissances transmises en fonction de différentes méthodes d'apprentissage:

- **apprentissage par coeur:**
le travail du sujet consiste essentiellement en un travail de mémorisation; le sujet ne connaît que ce qu'il a appris; le type de connaissances apprises est généralement un ensemble de faits ou de concepts;
- **apprentissage par l'exemple:**
le travail du sujet consiste à produire des connaissances générales à partir d'exemples ou de contre-exemples du domaine à acquérir; le sujet apprend à déduire des concepts généraux à partir de faits particuliers; le type de connaissances apprises est une habileté à résoudre un certain genre de problèmes spécifiques;
- **apprentissage par la découverte:**
la découverte implique la notion d'expérience, ou de simulation; le sujet se trouve placé devant une situation où il ne connaît pas forcément quels sont les concepts à découvrir; il est soumis à un travail actif d'observation, de recherche; le type de connaissances apprises dépasse largement la connaissance factuelle pour atteindre la compréhension et l'assimilation de principes ou de règles; le sujet se forge un mode de raisonnement, une façon d'aborder les problèmes et de les résoudre.

A priori, toute méthode d'apprentissage peut être valable. Toutefois, la qualité de celle-ci ne peut être déterminée qu'en fonction de nombreux autres facteurs, relatifs notamment aux caractéristiques propres à l'élève et au tuteur.

Le problème que nous nous posons dans cette thèse est de fournir une réflexion sur l'établissement d'une forme d'apprentissage adaptée à un certain domaine, pour un groupe de personnes défini. Ceci va nous permettre d'aboutir à la conception d'un système d'enseignement qui favorise l'apprentissage ainsi défini.

1.2 L'outil informatique en situation de formation

1.2.1 Un peu d'histoire

Le concept d'utilisation d'une machine en vue d'enseigner a commencé aux Etats-Unis durant les années 1920 grâce aux travaux du psychologue Sidney Pressey. Pressey a développé un concept présentant des questions à choix multiples et conservant la trace des réponses correctes et incorrectes. Les principes qu'il a développés ont été à la base du développement d'instructions programmées dans les années 50.

L'instruction programmée, ou enseignement programmé, est une méthode pédagogique systématique qui n'a, sous sa forme première, pas de rapport direct avec l'ordinateur. Selon une définition donnée par Houziaux [21], "un programme didactique peut se définir comme la suite des instructions que devra effectuer le couple élève - système d'enseignement". Dans un cadre scolaire traditionnel, le système d'enseignement représente un être humain. Le programme désigne alors simplement un relevé de matières à enseigner durant une période donnée. En revanche, dans le cadre de l'enseignement programmé, le système d'enseignement peut être par exemple un "livre programmé", ou une "machine à enseigner".

A la fin des années 50, il était possible d'identifier deux approches de base de l'instruction programmée:

1. la première approche souvent associée à Skinner [48] consiste en l'organisation des opérations dans une séquence linéaire. Chaque étudiant doit alors suivre la même série d'opérations dans le même ordre;
2. la seconde approche est basée sur une structure hiérarchique permettant aux étudiants de suivre différents chemins au travers du domaine. Un des travaux les plus importants effectués selon cette approche fut l'oeuvre de Crowder [9,10].

Les efforts d'informatisation des instructions programmées ont commencé sérieusement vers 1960. C'est en effet dans les années 60 que les premières expériences d'utilisation de l'ordinateur ont montré leur valeur potentielle en tant qu'outil d'éducation (ou de formation).

La notion de programme a alors évolué. Selon la définition de Houziaux [21],

il a alors pris le sens de "suite (ou ensemble organisé) d'instructions à faire exécuter par un ordinateur pour résoudre un problème déterminé".

Pour la plupart, les projets d'Enseignement Assisté par Ordinateur ont été développés de manière mutuellement exclusive: les universités, écoles et organisations de recherche ont trouvé de manière indépendante des moyens et des ressources pour construire des programmes individuels. De tels programmes ont été développés sur des ordinateurs différents, avec des langages de programmation différents, mais ils ont souvent reflété une philosophie d'éducation identique. Aucun de ces projets n'est toutefois devenu un standard de l'Enseignement Assisté par Ordinateur.

Dans les années 60, les ordinateurs étaient toutefois encore relativement coûteux, difficiles à utiliser et non familiers pour la plupart des enseignants.

Avec l'avènement du micro-ordinateur vers 1975, les développements sont devenus beaucoup plus nombreux. L'ordinateur a alors bénéficié de plusieurs avantages tels que:

- la portabilité;
- le coût initial faible;
- une relative facilité d'utilisation;
- l'autonomie d'utilisation.

De nos jours, les ordinateurs jouent un rôle toujours plus actif dans le processus d'enseignement. Ils deviennent familiers dans une salle de classe et leur demande augmente à tous les niveaux de l'éducation. Bien que l'Enseignement Assisté par Ordinateur ait fait l'objet d'un fort engouement ces dernières années, on peut toutefois constater que les programmes sont souvent développés à un niveau local, et qu'une large diffusion et la publication de ces programmes n'est pas encore de règle.

1.2.2 Avantages apportés par l'ordinateur

Le plus grand avantage apporté par l'ordinateur dans le domaine de l'enseignement est d'offrir un potentiel d'adaptation de la formation au niveau de l'étudiant:

- le flux de l'information, généralement imposé par l'organisation des chapitres dans un livre, peut être adapté de manière dynamique aux besoins individuels de l'étudiant, et par l'étudiant lui-même;
- ce qui est plus important encore, le rythme d'acquisition de la connaissance ne dépend plus de la personne qui dispense l'enseignement, mais de l'étudiant lui-même qui a alors la possibilité de passer davantage de temps dans l'apprentissage des domaines pour lesquels il a de la peine, puis plus rapidement dans l'apprentissage des domaines qu'il assimile plus facilement.

En outre, l'ordinateur apporte également des avantages de par ses caractéristiques propres:

- **mémoire:**
la mémoire permet le stockage des données, avec l'avantage de la taille des données stockées, mais également de la "stabilité" des données stockées, dans le sens où l'ordinateur ne se trompe pas, l'information fournie est toujours de la même rigueur, de la même qualité;
- **calcul:**
la capacité et la rapidité de calcul permet d'effectuer des exercices, de résoudre des problèmes, de répondre de manière pratiquement instantanée à toute demande qui nécessiterait plusieurs minutes, voire plusieurs heures par une résolution manuelle;
- **interactivité:**
l'ordinateur offre la possibilité d'un dialogue aisé entre la machine et l'apprenant, de même qu'un accès rapide aux informations.

L'ordinateur est encore dans la phase ascendante de sa diffusion et de son utilisation. Cet aspect de nouveauté peut également être considéré comme un avantage dans un domaine comme l'enseignement, où le travail de l'enseignant ne consiste pas uniquement à transmettre son savoir, mais également à motiver l'apprentissage, à le rendre attractif, captivant.

Il est important de remarquer également qu'un système d'apprentissage, quel qu'il soit, n'est bon pédagogiquement que s'il laisse une possibilité d'adaptation. Un bon système ne doit pas être figé, mais laisser une ouverture vers le futur. Dans cette optique, un des plus grands avantages des outils d'apprentissage électroniques réside dans leur potentiel d'adaptation du contenu et de la forme aux besoins de l'étudiant individuel.

1.2.3 Réticences vis-à-vis de l'ordinateur

Une certaine réticence vis-à-vis de l'utilisation de l'ordinateur découle parfois d'un sentiment de "déshumanisation" de l'apprentissage. En effet, si l'on envisage par exemple un cours en mathématiques, certains s'exclameront qu'une machine ne remplacera jamais un bon professeur de mathématiques pour en dispenser l'enseignement!

Une telle réaction n'est absolument pas dénuée de sens. Il est bien évident que, même si l'outil informatique est utilisé dans le processus de transmission ou d'acquisition de la connaissance, l'homme, le professeur ou l'expert occupe toujours une place prépondérante et fondamentale, notamment dans la phase de construction du système d'Enseignement Assisté par Ordinateur. En effet, il ne faut pas oublier que l'ordinateur n'est que le média utilisé pour transmettre la connaissance de l'expert à l'étudiant, au même titre que le serait un tableau noir ou un rétroprojecteur. Le siège de la connaissance n'est pas la machine. Celle-ci ne fait que transmettre le savoir de l'homme.

C'est dans la phase d'exploitation du système d'Enseignement Assisté par Ordinateur que le rôle de l'enseignant se modifie. Il est toutefois impossible de décrire ce rôle dans un cadre tout-à-fait général, d'une part à cause de la diversité des systèmes existants et du potentiel d'évolution dans ce domaine offert par le développement de l'informatique, et d'autre part car l'enseignant reste toujours maître devant sa classe. C'est donc lui qui décide de l'utilisation des possibilités nouvelles qui lui sont offertes.

Les problèmes liés à l'acceptation ou au rejet de l'utilisation de cette technologie nouvelle, de la part des enseignants, des étudiants, voire même de la part des parents, pourraient faire en soi l'objet d'une analyse sociologique approfondie. Nous n'entrerons toutefois pas plus profondément dans ce genre de considérations dans le cadre de cette thèse.

Le fait qui nous intéresse davantage est de constater que la façon de voir et d'aborder un cours a changé au cours du temps. Les écoles entreprennent des réformes, et la demande la plus forte émane des étudiants eux-mêmes qui suivent le rythme de la société qui les entoure.

1.3 Richesse d'un système de Formation Assistée par Ordinateur

L'Enseignement Assisté par Ordinateur constitue une aide dans un processus d'enseignement. Comme nous l'avons déterminé précédemment, la notion d'enseignement met en relation deux parties: l'enseignant et l'étudiant. Nous parlerons, dans le cadre de cette thèse, de préférence de **Formation Assistée par Ordinateur** dans le but d'élargir cette notion d'enseignement à une notion moins restrictive, qui inclut non seulement la formation dispensée par un enseignant à un étudiant, mais également la formation individuelle ou auto-formation.

L'Enseignement Assisté par Ordinateur est trop souvent assimilé à un simple **contrôle** de connaissances. Nous envisageons le problème sous un angle plus large, n'impliquant pas uniquement la vérification des connaissances acquises (par une suite d'épreuves ou de tests), mais également la **transmission** de connaissances, quel que soit le genre de connaissances transmises (théoriques ou pratiques).

Les deux principaux piliers d'un système de Formation Assistée par Ordinateur sont l'**individualisation** et l'**interactivité** de l'apprentissage.

1. Individualisation

L'individualisation, ou personnalisation de l'apprentissage est la justification fondamentale de tout système de Formation Assistée par Ordinateur.

Un tel système place la personne (utilisateur) dans une situation d'auto-formation, situation où elle organise elle-même son apprentissage.

Dans cette optique, l'outil de formation doit être utilisable et efficace par lui-même: il n'est pas un "gendarme" de la formation, mais plutôt un guide dans le processus d'acquisition des connaissances.

La notion d'individualisation implique également celle de souplesse d'utilisation:

- souplesse de **choix**:
le maximum de liberté est laissé à l'utilisateur dans le parcours de sa formation. Cette souplesse varie en fonction du type d'enseignement dispensé de même que du type de public visé: la liberté de

choix ne sera pas équivalente si l'objectif est l'enseignement d'un domaine où les connaissances à transmettre sont totalement nouvelles et indispensables à la suite de la formation que si l'objectif est le perfectionnement ou la vérification de connaissances supposées connues, ou partiellement connues;

- souplesse de **temps**, ou de **moment**:
la machine, contrairement au tuteur humain, n'est pas soumise à des limites de disponibilité ou de fatigue. Elle est accessible à tout moment, et toujours avec la même efficacité, la même précision;
- souplesse de **rythme de travail**:
l'utilisation de l'outil informatique a l'avantage de permettre un apprentissage adapté à la vitesse de lecture, de compréhension et d'assimilation de l'utilisateur.

2. Interactivité

L'interactivité est le second pilier sur lequel est fondée toute la technologie de la Formation Assistée par Ordinateur. En effet, l'ordinateur est sans doute, parmi les médias, celui qui autorise la plus grande interactivité.

L'interactivité doit être comprise au sens de processus de "dialogue" entre l'utilisateur et la machine. Celle-ci n'implique pas uniquement la notion de test ou d'épreuve où l'étudiant répond à des questions posées par la machine, mais également toute autre forme de communication ou d'interaction, que ce soit l'appel d'un index, d'une fonction d'aide, d'une fonction de calcul, etc.

Une combinaison judicieuse de ces deux atouts fondamentaux de la Formation Assistée par Ordinateur a pour avantages:

- de placer l'utilisateur dans des circonstances favorables pour qu'il puisse vérifier sa propre compréhension;
- de faire prendre conscience de la complexité de la réalité.

En effet, les premiers exercices effectués après l'apprentissage d'une notion peuvent être d'une simplicité telle qu'il est même parfois plus rapide de les résoudre manuellement.

Toutefois, la réalité n'est pas toujours aussi simple. L'objectif est également de donner des exercices qui ne pourraient pas, ou très difficilement

être résolus manuellement pour des raisons de complexité, ou simplement car le temps de résolution manuelle serait tellement long et fastidieux que des erreurs de calcul seraient pratiquement inévitables;

- d'inciter et aider à organiser des connaissances éparées en faisant accéder aux cadres, aux méthodes, aux outils qui situent ces connaissances et les mettent en relation.

Dans ce but, le logiciel doit donc allier la **découverte** à la **manipulation**.

- la découverte est un voyage à travers la base de connaissances du domaine étudié.
- la manipulation se fait par l'intermédiaire de programmes d'application des méthodes permettant par exemple de jouer avec les données, de visionner les résultats d'applications et des explications concernant les processus de résolution.

La combinaison de ces deux aspects permet d'aboutir à une plus grande autonomie de l'utilisateur.

1.4 **Connaissances nécessaires à la construction d'un système de Formation Assistée par Ordinateur**

La méthodologie de construction d'un système de Formation Assistée par Ordinateur est constamment confrontée à un problème de multidisciplinarité.

En effet, comme le montre la figure 1.2, le système final doit être le fruit d'une conjugaison harmonieuse entre l'**aspect pédagogique** de transmission des connaissances, l'**aspect scientifique** du savoir véhiculé, et l'**aspect technique** du mode de présentation des connaissances et du dialogue.

Chacun de ces aspects fait appel à des domaines de connaissance différents qu'il est nécessaire de maîtriser afin de créer un système équilibré et efficace.

- **Connaissance pédagogique**

L'objectif du système étant de répondre aux différents besoins de formation, l'aspect pédagogique tient une place fondamentale dans sa conception.

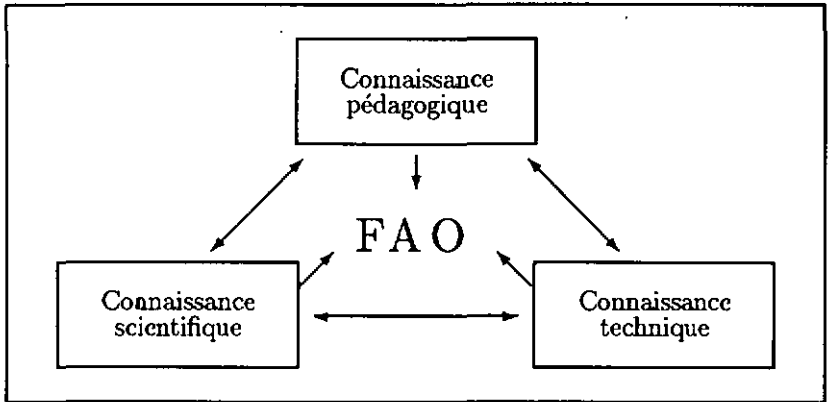


Figure 1.2: Types de connaissances nécessaires à la construction d'un système de Formation Assistée par Ordinateur (FAO)

Les questions auxquelles nous sommes confrontés sont par exemple:

- A quel type de public s'adresse-t-on?
- Comment faire passer un message de la manière la plus attractive possible?
- Comment faut-il structurer la présentation de l'information sans entrer en conflit avec le niveau de perception de l'utilisateur?
- De quelle manière l'apprentissage peut-il être envisagé sous forme de jeu?
- Quels moyens utiliser pour susciter l'envie d'apprendre, de progresser, d'aller voir plus loin?
- Comment attirer l'attention de l'utilisateur sur les points essentiels, ou plus simplement sur les informations que l'on veut transmettre?

Toutes ces questions méritent d'être étudiées en vue d'aboutir à une méthodologie de formation solide et cohérente.

• Connaissance scientifique

La connaissance scientifique est une connaissance de la matière à transmettre, du contenu à intégrer dans le système de formation.

Une étude approfondie du contenu permet de structurer l'information en vue de faciliter son apprentissage, de le clarifier.

Une bonne connaissance du domaine à enseigner doit permettre de répondre à des questions du type:

- Quel niveau d'apprentissage voulons-nous transmettre?
- Quelle étendue du contenu voulons-nous traiter?
- Comment organiser nos connaissances, compte tenu de la manière dont ces connaissances vont être transmises?

● **Connaissance technique**

L'informatique doit être conçue comme un **moyen**, un outil de travail permettant d'atteindre des objectifs d'attractivité, de flexibilité, d'interactivité, etc. Comme le soulèvent Besnainou et al [5], "C'est à l'ordinateur de servir la pédagogie, et non l'inverse".

Bien que l'informatique soit un outil au service de l'objectif poursuivi, il est toutefois fondamental d'y accorder une attention toute particulière. En effet, la qualité de la formation transmise ne dépend pas uniquement de la qualité des connaissances véhiculées, mais également - et même très fortement - du support de transmission de ces connaissances.

La maîtrise de l'outil informatique doit alors permettre de répondre à des questions techniques de construction du système:

- Comment organiser la base de données dans le but d'optimiser les temps d'accès?
- De quelle manière concevoir le générateur de la base de données en vue de faciliter la saisie des données?
- Quels périphériques d'entrée et de sortie faut-il prévoir?
- Comment améliorer l'interactivité et la convivialité du système?

Toutes ces questions ne sont bien évidemment qu'un aperçu de l'ensemble des étapes menant de la conception à la réalisation du logiciel. Des éléments de réponse à ces différentes questions seront amenés tout au long de l'ouvrage.

1.5 Processus de production d'un système de Formation Assistée par Ordinateur

Le processus de production d'un système de Formation Assistée par Ordinateur peut être considéré comme une suite d'étapes, où l'on trouve en amont la naissance d'une idée, et en aval un produit concret et directement utilisable.

Si l'on considère ces différentes étapes d'un point de vue suffisamment général, le processus ne dépend pas du type de logiciel développé. Sommerville [50] a proposé une représentation schématique des principales étapes du processus de développement d'un logiciel.

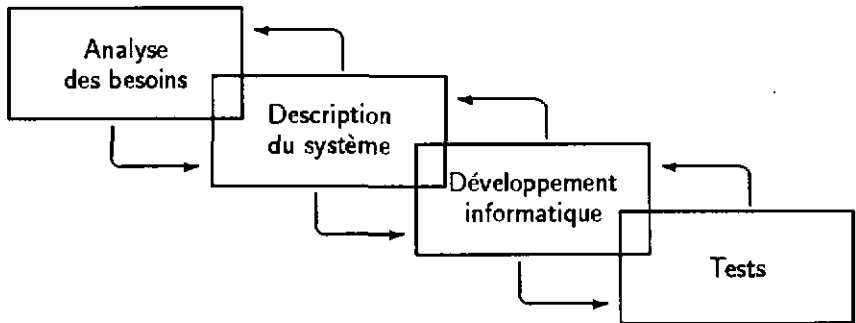


Figure 1.3: Processus de production d'un logiciel (Source: Sommerville I. (1985). *Software Engineering*, second edition. Addison-Wesley Publishing Company)

Comme on peut le voir sur la figure 1.3, ce processus est composé de quatre étapes fondamentales.

1. Analyse des besoins

L'analyse des besoins est la première étape du processus. C'est une étape fondamentale qui a pour but de déterminer les objectifs auxquels le logiciel devra répondre, sur la base d'une étude approfondie de la population des futurs utilisateurs, de leurs attentes et de leurs besoins.

2. Description du système

Sur la base des objectifs définis dans la première étape, il s'agit de déterminer quels sont les moyens (matériels et logiciels) à mettre en oeuvre afin d'atteindre ces objectifs. Dans le cadre de la création d'un système de Formation Assistée par Ordinateur, il s'agit également de structurer les connaissances à transmettre.

Cette étape doit aboutir à une description détaillée du système. Elle permet de vérifier la cohérence et la pertinence des choix effectués dans l'étape précédente, afin d'y retourner le cas échéant.

3. Développement informatique

Cette troisième étape est la phase de réalisation effective du logiciel. Elle consiste à adapter l'ensemble du projet à la spécificité du média utilisé, en l'occurrence à l'ordinateur. A cette étape, chaque programme est testé individuellement.

4. Tests

Tous les programmes sont intégrés et testés en tant que système complet en vue de s'assurer que les objectifs ont bien été atteints.

Chacune de ces quatre étapes comprend une phase de contrôle qui permet de faire le point sur le travail réalisé, le processus ne pouvant se poursuivre que lorsque l'étape précédente est parfaitement approuvée.

Sommerville [50] met ici l'accent sur la distinction entre deux activités de contrôle différentes, mais nécessaires toutes deux à chaque étape:

- la **vérification** qui doit répondre à la question "construisons-nous le produit correctement?"
- et la **validation** qui doit répondre à la question "construisons-nous le bon produit?"

Ces deux types de contrôle doivent permettre d'identifier les éventuelles erreurs, et forcer le cas échéant le retour à une des étapes précédentes.

La distinction entre ces deux types de contrôle est généralement admise, mais dans une terminologie parfois différente. De Montmollin [36] parle des **expérimentations** qui ont lieu "en cours de la fabrication du programme" et permettent de vérifier les erreurs dans la construction du produit. La **validation**, considérée comme la phase terminale du processus, est "la mesure de l'efficacité réelle, selon les critères les plus objectifs possibles". Elle a pour but de mesurer la valeur du programme dans une situation normale d'emploi, qui peut fort bien différer de la situation expérimentale.

Quelle que soit la terminologie utilisée, l'idée principale à retenir est que le contrôle est une activité fondamentale du processus. Il est indispensable à chaque étape, et constitue un gage de la qualité du travail réalisé.

Nous n'aborderons toutefois pas les problèmes de tests, de contrôle, de corrections et d'amélioration des programmes dans le cadre de cette thèse. Nous

nous concentrerons sur une étude détaillée des trois premières étapes du processus, dans le cas particulier de la création d'un logiciel de Formation Assistée par Ordinateur.

1.6 Spécificités du type d'application

L'établissement d'un cours d'enseignement d'une matière quelconque peut être soutenu par certaines règles générales, essentiellement de nature pédagogique. Toutefois, celles-ci ne peuvent que déterminer les grandes lignes directrices de l'établissement du cours. Il convient ensuite de considérer les spécificités de la matière à enseigner afin de déterminer de façon détaillée l'organisation du cours.

Il en est de même pour la construction d'un système de Formation Assistée par Ordinateur: certains principes de base régissent les aspects généraux de la construction du système. Ceux-ci concernent notamment: des principes ergonomiques qui visent à concevoir une application bien adaptée aux conditions de travail de l'utilisateur; des principes pédagogiques qui visent à améliorer le processus d'éducation ou de formation; des principes techniques qui visent à utiliser au maximum les possibilités offertes par le matériel informatique.

La prise en considération de ces règles générales permet de déterminer "ce qu'il faudrait faire", "ce qu'il vaut mieux ne pas faire", "ce qu'il est possible de faire", en fonction de l'avancement technologique et du niveau de connaissances pédagogiques. Elle ne permet toutefois pas d'aboutir à un produit fini, puisque la structure de chaque application dépend inévitablement de la matière à enseigner.

Nous nous intéressons dans le cadre de cette thèse à la formation dans un domaine bien spécifique qui est celui de la **statistique**.

La statistique est un domaine particulièrement bien adapté à notre problème: elle est une science en soi, une science qui détermine tout un domaine de recherche, mais c'est également une science appliquée.

Enseigner la statistique, c'est enseigner des techniques d'utilisation de méthodes, d'outils, c'est faire acquérir un savoir, une connaissance des moyens que l'homme a à sa disposition pour analyser, étudier, décortiquer un problème. Mais c'est également faire acquérir un savoir-faire: une méthode peut être comprise, acquise, maniée avec dextérité. Mais l'utilisation de cette méthode

est-elle adéquate en fonction du problème à résoudre?

Une moyenne, par exemple, est une notion que l'on connaît depuis notre plus jeune âge. C'est une notion qui nous paraît tout-à-fait naturelle. Nous avons une quantité d'information et il nous paraît naturel de synthétiser l'ensemble de nos informations en en calculant une moyenne. Mais telle moyenne, calculée sur tel ensemble de données, a-t-elle un sens? Très souvent, nous nous contentons d'appliquer les méthodes, d'obtenir des résultats, et de les interpréter sans connaître réellement la validité des résultats que nous analysons.

L'enseignement de la statistique ne doit donc pas se contenter de présenter les bases mathématiques du développement de tous les outils et de toutes les méthodes, mais également encourager l'acquisition d'un savoir pratique. Ceci est un problème fondamental, car l'application des méthodes statistiques est très souvent longue et fastidieuse. Illustrons ce problème par un exemple: imaginons que lors d'un cours, on désire montrer aux étudiants l'effet d'une valeur aberrante sur la moyenne arithmétique et la médiane. Sur la base de 10 chiffres,

5 5.5 4 3.5 4 4.5 5 5.5 3.5 6

on leur demande de calculer la moyenne et la médiane. On modifie ensuite la valeur d'un des chiffres (considérons par exemple qu'une erreur de saisie a modifié la valeur du 4.5 en une valeur aberrante de 45). On leur demande alors de recalculer la moyenne et la médiane sur ce nouvel ensemble d'observations pour ensuite discuter des résultats obtenus. L'ensemble de ces opérations étant relativement simple et rapide, l'intégration d'un tel exemple dans un cours ne posera sans doute pas de grands problèmes.

Mais imaginons maintenant que l'on veuille faire la même démonstration non plus sur la moyenne arithmétique, mais sur une droite de régression, sur une analyse de variance ou sur le résultat d'un test d'hypothèse. Le temps matériellement nécessaire à la réalisation de tels exemples rend pratiquement impossible leur intégration dans un cours. Or, l'apprentissage véritable ne se fait que lorsque l'on s'est "mouillé les mains", lorsque l'on a fait une multitude d'essais, que l'on a pu observer par soi-même le résultat de telle analyse sur telles données.

Notre objectif est d'apporter une solution à ce genre de problèmes en élaborant une méthode permettant de transmettre non seulement le savoir, mais également, et surtout le savoir-faire:

- le savoir par l'apport d'une connaissance théorique du domaine;

- le **savoir-faire** par l'acquisition d'une **expérience pratique**.

Tout au long de notre exposé, nous illustrerons nos propos par le cas pratique du développement d'un système de formation en statistique. Ceci nous permettra de ne pas rester à la présentation de principes généraux, mais au contraire d'apporter des exemples concrets d'application de ces principes.

Partie I

**CONCEPTION
PEDAGOGIQUE**

Chapitre 2

Analyse pédagogique

2.1 Introduction

Notre objectif est de développer un système de Formation Assistée par Ordinateur. Mais quel genre de système désirons-nous développer? Quels sont les besoins spécifiques auxquels nous voulons répondre? A quel type de personnes notre application va-t-elle s'adresser? Quelles connaissances voulons-nous transmettre?

Ce premier chapitre a pour but de fournir des réponses à ces différentes questions afin d'établir une description précise des objectifs que nous voulons atteindre par le développement de notre application. Nous aboutirons à une première description globale de l'application après avoir déterminé:

- les références disponibles dans le domaine d'apprentissage considéré;

- le public potentiellement touché par l'application;
- les types de besoin auxquels celle-ci est destinée à répondre;
- les types d'apprentissage qui en découlent;
- la nature et l'étendue des connaissances à transmettre.

Les principaux objectifs ayant été déterminés, nous serons alors en mesure d'expliquer la stratégie pédagogique soutenant l'application.

2.2 Types de références disponibles

Avant de prétendre définir une nouvelle méthode d'enseignement, il est nécessaire d'étudier les références existantes, afin d'en dégager les différents points forts et points faibles. Ceux-ci permettront de mettre en évidence les éventuelles insuffisances de ces méthodes, afin de poser les premières bases de notre système.

Dans le domaine spécifique de la statistique, sur lequel nous nous baserons par la suite afin d'illustrer nos propos par des exemples concrets, nous avons pu déterminer trois catégories de références.

1. Ouvrages classiques de statistique

Il existe d'innombrables ouvrages de statistique, que ce soit des livres de statistique élémentaire ou des livres traitant de domaines très spécifiques.

Toutefois, nous pouvons remarquer que de tels ouvrages ne répondent pas forcément aux besoins de l'ensemble du public potentiel.

Pour un débutant dans le domaine, l'apprentissage par la lecture de livres est souvent long et fastidieux. Il nécessite une attention soutenue qui est d'autant plus pénible que le lecteur est passif devant son livre.

Les demandes en information d'un utilisateur de méthodes statistiques sont souvent ponctuelles et très spécifiques. Les ouvrages de base ne répondent pas adéquatement à ce genre de besoin car l'information spécifique désirée est souvent difficile à trouver. Les ouvrages spécialisés quant à eux sont très souvent d'une complexité qui n'est accessible que par la minorité des "statisticiens professionnels", ou des chercheurs en statistique.

2. Dictionnaires et encyclopédies

En dehors des ouvrages de statistique classiques, nous trouvons également des dictionnaires, ou encyclopédies en statistique.

Ceux-ci sont pour la plupart en langue anglaise, ce qui restreint déjà le public visé.

Parmi cette catégorie de références, nous trouvons:

- d'une part des dictionnaires, tel que le *Dictionary of Statistical Terms* publié par l'Institut International de Statistique [31]. Le but d'un tel ouvrage est de fournir une explication concernant les termes statistiques les plus couramment utilisés. Leur définition est décrite de manière très simple et claire mais également très concise, et les méthodes n'y sont pas développées;
- d'autre part des encyclopédies traitant de l'ensemble des domaines statistiques. Nous pouvons notamment citer dans cette catégorie deux encyclopédies diffusées mondialement: l'*Encyclopedia of Statistical Sciences* en neuf volumes de S. Kotz et N.L. Johnson [25] et l'*International Encyclopedia of Statistics* de W.H. Kruskal et J.M. Tanur [27] où l'information fournie est très vaste. Toutefois, la complexité de telles encyclopédies dépasse très souvent l'entendement du profane.

3. Logiciels statistiques

Les logiciels statistiques offerts sur le marché sont très nombreux. Nous trouvons notamment des logiciels tels que SAS, SPSS, BMDP, MINITAB, GAUSS qui permettent de manière relativement aisée d'obtenir des analyses complètes sans connaissances préalables trop élevées. La simplicité d'utilisation de ces logiciels peut représenter un danger puisque leur utilisation n'implique pas nécessairement la compréhension des outils ou des méthodes employées.

Nous trouvons également un grand foisonnement de logiciels permettant de réaliser certaines tâches statistiques spécifiques. Souvent toutefois, ces logiciels spécialisés ne peuvent être utilisés adéquatement que par des experts du domaine.

Le premier commentaire que nous pouvons faire est que le nombre de références est énorme. Entre les ouvrages élémentaires, les ouvrages spécialisés, les encyclopédies et les logiciels offerts sur le marché, il peut sembler qu'il doit

obligatoirement exister une source d'information permettant de répondre à n'importe lequel de nos besoins.

Ceci n'est toutefois pas vrai pour tout type de public considéré. Si les références existantes sont généralement excellentes pour le statisticien proprement dit, il n'en est pas de même pour un public large à un niveau plus élémentaire. Face à l'ensemble de ces références, nous pouvons remarquer qu'il existe un manque. Et ce manque, c'est un lien entre la théorie véhiculée par les ouvrages et encyclopédies en statistique, et l'utilisation des méthodes statistiques proposée par les logiciels.

L'approche traditionnelle offerte par les différents ouvrages ou dictionnaires en statistique est purement statique. Quant aux logiciels, s'ils fournissent (par une aide interactive ou par le manuel d'utilisation) des explications détaillées concernant l'utilisation des routines proposées, ils ne permettent que rarement d'obtenir des explications simples concernant les méthodes elles-mêmes, leurs champs d'application, des exemples illustrant leur fonctionnement.

Nous nous proposons de tenter de combler ce manque par un système d'apprentissage qui permette de lier ces deux aspects indissociables dans la compréhension de la statistique: la théorie et la pratique.

2.3 Public potentiel

Bien qu'il nous semble souvent connaître intuitivement le public auquel on s'adresse, il faut bien se rendre compte que si les **caractéristiques générales** de ce public sont relativement aisées à fournir (élèves d'une école secondaire, étudiants dans un certain domaine dans une école supérieure, etc), il est par contre très difficile de déterminer les **caractéristiques personnelles** des individus formant cette population.

Les caractéristiques générales du groupe sont importantes car elles peuvent servir de fil conducteur dans l'élaboration de l'application, mais il ne faut pas perdre de vue qu'un des fondements de la Formation Assistée par Ordinateur est l'**individualisation**, la personnalisation de l'apprentissage. Il est donc important d'effectuer une analyse un peu plus profonde des différents facteurs caractéristiques de cette population.

Besnainou et al [5] proposent un ensemble de facteurs personnels et situationnels pouvant servir de base à la caractérisation de la population cible. En voici quelques exemples:

1. Facteurs personnels

- la **structure cognitive**, ou organisation de l'ensemble des connaissances dont une personne dispose à un moment donné;
- le **stade de développement cognitif**, ou niveau de maturation caractéristique du stade de développement;
- la **capacité intellectuelle**, ou niveau général de l'intelligence;
- les **facteurs motivationnels**, ou ce qui influence le désir de faire quelque chose, d'atteindre un objectif.

2. Facteurs situationnels

- les **facteurs socio-psychologiques**, ou relation élève - enseignant;
- l'**environnement**, ou situation dans laquelle se déroulera l'application.

Bien que de tels facteurs peuvent fournir des éléments de réflexion, il est assez rare de pouvoir caractériser la population de manière très stricte. Il est bien plus fréquent de devoir faire face à de nombreuses inconnues, notamment en ce qui concerne les caractéristiques personnelles des individus. Si de telles caractéristiques sont inconnues ou trop vagues, il peut alors s'avérer nécessaire de fixer des pré-requis.

Examinons les caractéristiques générales du public cible pour notre application particulière:

Notre objectif est de créer un "pont" dynamique entre la littérature classique et les logiciels spécialisés en statistique. Celui-ci n'est donc pas destiné à une classe privilégiée, sous-entendu les professionnels du domaine, mais plutôt à un large spectre de la population, allant des étudiants du niveau gymnasial à l'université, aux chercheurs qui utilisent occasionnellement la statistique comme outil de démonstration en passant par les cadres qui peuvent utiliser ce logiciel soit comme aide au développement de leurs idées, soit pour la création de graphes de base ne nécessitant pas l'emploi de logiciels spécialisés.

Nous pouvons dégager trois groupes fondamentaux de personnes potentiellement touchées par un problème de recherche d'information (ou de besoin de

formation). Notre critère de classification est la capacité à trouver toute information spécifique ou globale recherchée dans le domaine traité.

1. Groupe “novice”

Appartient à ce groupe toute personne ne possédant aucune connaissance, pratique ou théorique, dans un domaine particulier - ou dans tous les domaines - de la statistique.

2. Groupe “utilisateur occasionnel”

Le profil de l'utilisateur occasionnel peut être très variable. Il peut correspondre

- au chercheur utilisant la statistique comme outil d'analyse. Ses demandes d'information sont ponctuelles et variées;
- à l'étudiant dans tout domaine touchant de près ou de loin au domaine de la statistique: les sciences (physique, chimie, biologie, mathématiques); les sciences politiques, économiques ou sociales; la médecine; etc;
- au lecteur occasionnel d'articles statistiques.

3. Groupe “professionnel”

Ce groupe constitue le groupe des statisticiens proprement dits, maîtrisant les techniques et les méthodes statistiques dans leur ensemble, de même que leur utilisation.

Remarquons que cette classification, que nous avons illustrée pour le cas spécifique d'un apprentissage dans le domaine de la statistique, pourrait également être valable pour de nombreux autres domaines.

Nous nous intéressons dans le cadre de notre application aux groupes “novice” et “utilisateur occasionnel” car ce sont les deux groupes les plus intéressants au niveau de leurs besoins de formation. Le groupe “professionnel” ne nous intéresse pas ici, le niveau de formation de même que les centres d'intérêt et la complexité des domaines traités étant très différents pour ce groupe que pour les deux précédents.

En fonction des caractéristiques générales définies ci-dessus, le spectre du public potentiellement touché semble être relativement large. L'information que nous pouvons donner concernant les différents facteurs proposés par Besnainou

et al [5] ne peut donc être qu'indicative, voire même assez floue:

1. Facteurs personnels

- Structure cognitive: l'organisation de l'ensemble des connaissances dont dispose une personne au moment de l'utilisation du logiciel est a priori inconnu.
- Stade de développement cognitif: notre application étant destinée à un public adulte (ou d'étudiants avancés), il nous est impossible de déterminer le stade de développement cognitif des individus qui le composent. Nous supposons un stade de développement cognitif relativement élevé vu que notre application touche un public adulte ayant une formation préalable de niveau supérieur.
- Capacité intellectuelle: le domaine couvert par notre application est d'un niveau d'enseignement universitaire. La capacité intellectuelle requise est donc la maîtrise, ou tout au moins la connaissance des principes mathématiques de base nécessaires à l'entrée dans une formation universitaire.
- Facteurs motivationnels: notre application est orientée vers un public à la recherche d'information. Nous ne cherchons pas à imposer un apprentissage "forcé", mais plutôt à répondre à une demande d'information.

2. Facteurs situationnels

- Facteurs socio-psychologiques: ces facteurs sont peu pertinents dans une situation d'auto-formation. Nous pouvons donc les négliger.
- Environnement: l'utilisation du logiciel peut se faire dans une très large variété de contextes (bureau, école, etc), d'où l'importance d'avoir une application très autonome (qui se comprend par elle-même).

2.4 Types de besoins

En fonction des deux groupes que nous avons retenus ("novice" et "utilisateur occasionnel") et des différentes caractéristiques que nous avons pu mettre en évidence, nous sommes en mesure de déterminer de manière plus précise les types de besoin auxquels l'application doit pouvoir faire face.

Une détermination précise des types de besoins est essentielle, dans la mesure où l'aptitude à répondre aux besoins des utilisateurs est le principal gage de la qualité d'un logiciel.

Les besoins de formation peuvent se situer à plusieurs niveaux.

1. **Besoin ponctuel**

Un besoin ponctuel correspond à une recherche d'information concernant certains aspects d'une notion connue mais non maîtrisée.

Toute personne - chercheur ou utilisateur de méthodes ou d'outils statistiques - possédant des connaissances de base dans le domaine traité et ayant des lacunes spécifiques qu'elle est capable d'identifier se trouve confrontée à un **besoin ponctuel** d'information.

Le besoin ponctuel est donc caractéristique du deuxième groupe de public potentiel, le groupe "utilisateur occasionnel".

2. **Besoin structuré**

Un besoin structuré correspond à une recherche d'information globale. C'est un besoin de pouvoir bénéficier d'une formation suivie et méthodique du domaine dans sa globalité ou d'un sous-domaine particulier.

Toute personne n'ayant aucune connaissance - ou des connaissances très précaires - du domaine en question et désirant bénéficier d'une formation continue et assistée est confrontée à un **besoin structuré** d'information.

Le besoin structuré est caractéristique du premier groupe de public potentiel, le groupe "novice".

Bien que chaque type de besoin ainsi défini soit caractéristique d'un des groupes préalablement déterminés, il est important de noter que ces deux types de besoins ne sont pas nécessairement incompatibles:

- un besoin structuré peut soulever un besoin ponctuel: généralement, l'apprentissage d'un domaine nécessite des connaissances préalables qui peuvent être supposées connues mais s'avérer non maîtrisées; d'où l'émergence d'un besoin ponctuel d'information;
- un besoin ponctuel peut soulever un besoin structuré: l'étude d'une information ponctuelle peut soulever un problème de compréhension de plus vaste envergure qui nécessitera d'approfondir tout un domaine; d'où l'émergence d'un besoin structuré d'information.

2.5 Types d'apprentissage

Les connaissances de base permettant de répondre à ces deux types de besoin sont les mêmes. Toutefois, la manière de les transmettre est différente.

A chaque type de besoin correspond donc un type d'apprentissage.

En vue de répondre à ces deux niveaux de besoin, nous allons présenter l'information de base sous la forme d'une **base de connaissance**. Comme l'illustre la figure 2.1, cette base de connaissance, ou **base de données**, peut être schématisée comme une pièce contenant des armoires, chaque armoire étant composée d'une multitude de tiroirs.

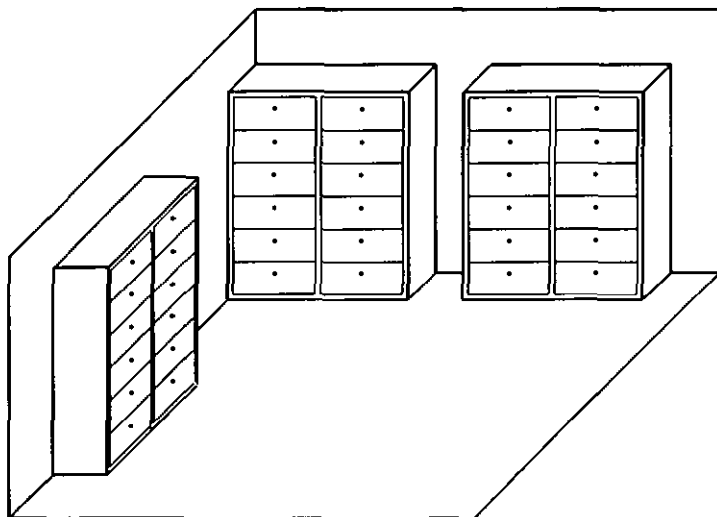


Figure 2.1: Représentation schématique de la base de connaissances

Dans la figure 2.1, la pièce représente l'ensemble de nos connaissances. Nous pouvons structurer nos connaissances. Si nous devons ranger des légumes et de la viande, nous ne mettrions certainement pas ces différents articles dans la même armoire. De la même manière, nous allons déterminer des sous-ensembles cohérents de nos connaissances que nous "rangerons" dans des armoires spécifiques. En vue de nous permettre de retrouver facilement l'information dont nous avons besoin, nous mettrons dans chaque tiroir une information précise concernant une notion particulière.

Cette structure doit nous permettre de répondre aux deux niveaux de besoins envisagés par deux types d'apprentissage.

1. Apprentissage ponctuel

La structure définie ci-dessus répond au besoin d'apprentissage ponctuel: une personne ayant un besoin bien défini peut aisément entrer dans la pièce, eboisir l'armoire dont il a besoin et ouvrir le tiroir correspondant à l'information qu'il désire obtenir.

2. Apprentissage structuré

L'apprentissage structuré est en quelque sorte une visite guidée de la pièce: en vue d'offrir la possibilité d'un apprentissage structuré, nous pouvons envisager de guider la personne dans son parcours de la base de connaissance. A tout moment, et pour autant qu'elle le désire, nous pourrions éventuellement lui faire passer un test nous permettant de nous rendre compte de l'état de ses connaissances et de lui indiquer le chemin à suivre pour combler ses lacunes et continuer son apprentissage.

2.6 Détermination du contenu

2.6.1 Le vaste domaine de la statistique

La statistique est une discipline dont les racines remontent à plusieurs siècles. Considérons par exemple la notion de moyenne arithmétique qui est une des plus anciennes méthodes utilisée pour combiner des observations afin d'obtenir une valeur approximative unique. Son utilisation semble en effet remonter au temps des astronomes babyloniens au troisième siècle avant J.-C, pour la détermination de la position du soleil, de la lune et des planètes.

La statistique en tant que discipline provient de l'étude et de la description des états, sous leurs aspects économiques et encore plus spécifiquement démographiques. En effet, des formes de recensement se pratiquaient déjà à l'époque des grandes civilisations de l'Antiquité, dont l'étendue et la complexité des empires nécessitaient une administration rigoureuse. Les pratiques ont bien entendu fortement évolué avec le temps, des méthodes nouvelles se sont développées, et, de la simple collecte des données, la statistique a petit à petit pris un sens nouveau. En plus de la notion de "comptage" ou de "mesure" est venu s'ajouter le problème de la comparaison. Or, la comparaison

de mesures nécessite une connaissance de leur exactitude, de leur précision, de la manière d'exprimer l'incertitude rattachée aux valeurs. C'est là où la statistique inférentielle et la théorie des probabilités viennent prendre toute leur importance.

On trouve dès le 17^{ème} siècle de grands savants qui ont participé au développement de la statistique telle qu'on la connaît aujourd'hui. Citons par exemple Jacques Bernoulli au 17^{ème} siècle; Pierre Simon Laplace, Thomas Bayes, Rudger Josif Boskovic au 18^{ème}; Adrien Marie Legendre, Carl Friedrich Gauss, Siméon Denis Poisson, au 19^{ème}; Karl Pearson, William Sealy Gosset, Ronald Aylmer Fisher, Jerzy Neyman, Egon Sharpe Pearson au 20^{ème}; et on pourrait en citer encore beaucoup d'autres. Tous ont contribué à faire de la statistique une science en soi.

La statistique moderne est beaucoup plus qu'un ensemble de méthodes, d'outils, ou de techniques isolées utilisées dans des sciences individuelles et spécifiques. Il existe une unité des méthodes statistiques, une logique cohérente, une base scientifique, mathématique et probabiliste, solide.

La statistique n'est véritablement reconnue comme discipline en tant que telle que depuis le début du 20^{ème} siècle.

C'est donc une science relativement jeune et encore en pleine expansion, en pleine recherche. De nombreuses branches nouvelles, un grand nombre de spécialisations et d'applications se sont développées: la théorie de la décision, les séries chronologiques, l'analyse multivariée, l'économétrie, la théorie des jeux, l'inférence non paramétrique. La statistique mathématique et ses applications continuent toujours de se développer, de s'étendre.

2.6.2 Etendue et nature des sources d'information retenues

Comme nous pouvons le constater, la statistique est un domaine très vaste, beaucoup trop vaste pour être synthétisé dans un seul ouvrage. Notre objectif n'est pas d'exposer les développements récents et les nouvelles méthodes d'analyse faisant encore intégralement partie du domaine de la recherche. Nous laissons ces domaines pointus aux ouvrages spécialisés.

Comme nous l'avons déterminé en section 2.3, nous nous intéressons davantage à toucher un public de "novices" ou d'"utilisateurs occasionnels". Notre objectif est donc de présenter les bases, les fondements de la statistique actuelle. Nous proposons d'inclure dans notre système de Formation Assistée

par Ordinateur les principaux aspects relatifs aux domaines de la statistique descriptive, de la théorie des probabilités et de la statistique inférentielle.

- **Statistique descriptive:** mesures de tendance centrale, de dispersion, d'asymétrie et d'aplatissement; présentation des données, sous forme de tableaux ou de représentation graphique; calcul d'indices; analyse de séries chronologiques; méthodes d'analyse de données.
- **Probabilités:** lois de probabilité discrètes et continues; théorèmes de convergence; analyse combinatoire.
- **Statistique inférentielle:** méthodes d'échantillonnage; méthodes d'estimation ponctuelle et par intervalle; tests d'hypothèses paramétriques et non paramétriques; analyse de variance; analyse de régression; plans d'expériences.

Les documents scientifiques de référence pour notre étude ont été choisis en fonction de la théorie que nous voulons transmettre, l'objectif n'étant pas de rendre des savants encore plus savants dans le domaine, mais plutôt de permettre au néophyte de se former, de manière **ponctuelle** ou **continue** en acquérant les bases théoriques des notions dans un contexte tourné vers la pratique, vers l'utilisation des notions acquises.

Notre principale base de travail a été l'*Encyclopédie et Dictionnaire en Statistique* édité par Dodge [12]. En effet, cet ouvrage permet de répondre aux différents besoins soulevés précédemment, autant par sa structure que par son contenu.

2.7 Description globale de l'application

L'analyse du public potentiel et de ses besoins nous a permis de déterminer les deux niveaux d'apprentissage auxquels notre application est destinée à répondre: l'apprentissage **ponctuel** et l'apprentissage **structuré**.

Le premier type d'apprentissage peut être assimilé à une recherche directe dans un dictionnaire, ou une encyclopédie.

Le second type d'apprentissage correspond davantage à la lecture séquentielle d'un ouvrage, ou d'un livre.

L'adaptation de ces deux types d'apprentissage au domaine de la Formation Assistée par Ordinateur nous amène à parler des notions de **dictionnaire électronique** et de **livre électronique**:

- le dictionnaire électronique a pour but d'offrir des informations spécifiques relatives à des mots ou des notions, ce qui permet un accès direct à toute information ponctuelle recherchée. De plus, l'information à l'intérieur de chaque mot sera structurée en thèmes, de telle sorte que l'on puisse accéder directement non seulement au mot désiré, mais également à une information spécifique à l'intérieur de chaque mot;
- le livre électronique, à l'image d'un livre classique, présente les différents domaines traités selon une structure que l'on peut qualifier de table des matières. Le texte est rédigé de telle manière que la lecture séquentielle de l'ensemble du texte soit logique, cohérente, fluide, permettant ainsi un apprentissage structuré.

Si le dictionnaire et le livre électroniques se contentaient de donner, sous quelque forme que ce soit, un accès à de l'information textuelle, notre objectif ne serait pas totalement rempli puisque nous ne véhiculerions que l'aspect **théorique** des domaines traités. Nous avons mis précédemment l'accent sur l'aspect **pratique** de l'apprentissage. Nous proposons d'apporter cet apprentissage pratique, ou apprentissage du savoir-faire, par une bibliothèque de programmes d'application que nous nommons **outils statistiques**.

Les outils statistiques doivent permettre à l'utilisateur (du livre ou du dictionnaire électronique) de se mettre lui-même à l'épreuve, de participer activement à la compréhension des notions acquises par le texte. Dans cette optique, nous proposons de mettre à sa disposition une "feuille de travail", dans laquelle il peut alors tester les méthodes sur ses propres données afin d'en observer non seulement les résultats, mais également les formules utilisées et des explications concernant les calculs effectués.

Le concept que nous proposons est donc un concept intégré, offrant en une seule application ces trois éléments fondamentaux: le dictionnaire électronique, le livre électronique et les outils statistiques.

La figure 2.2 offre une représentation schématique de l'articulation entre ces trois éléments. Les caractéristiques principales en sont les suivantes:

1. L'application possède deux entrées principales: par le livre électronique

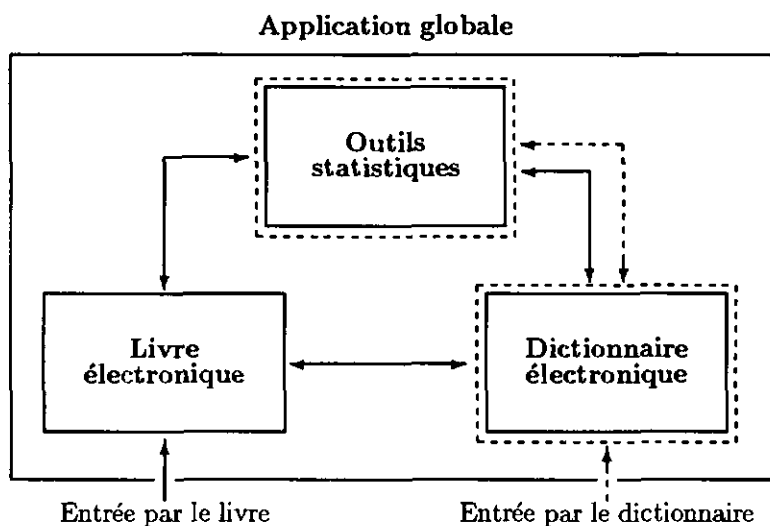


Figure 2.2: Structure globale de l'application statistique

(les possibilités liées à cette entrée sont représentées en trait plein) ou par le dictionnaire électronique (pointillé).

2. Le livre électronique donne un accès direct à la consultation du dictionnaire électronique.
3. Les outils statistiques sont accessibles par le dictionnaire électronique et par le livre électronique.

Nous n'avons présenté ici qu'un aperçu très général de ces trois éléments fondamentaux. Une description plus détaillée est proposée dans le chapitre 4.

2.8 Stratégie pédagogique

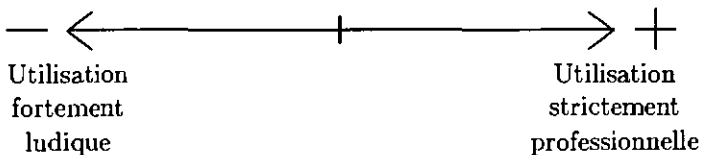
Les premières bases ayant été posées, nous sommes à présent en mesure de dégager les lignes fondamentales de la stratégie dans laquelle s'inscrit notre application.

Hennart et al. [19] proposent un modèle schématisé de repérage en vue de situer un logiciel en fonction de trois axes pédagogiques principaux.

Une stratégie pédagogique peut alors être définie par le positionnement de l'application, ou du système sur ces trois axes:

- **Axe 1: "ancrage du réel"**

L'axe "ancrage du réel" permet le positionnement du système par rapport à son utilisation. Il donne une indication concernant le **mode de transmission de l'information**.

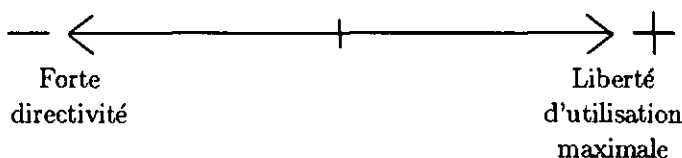


Une application d'utilisation strictement professionnelle offre des informations claires, précises, bien définies. Les critères principaux de présentation de l'information sont l'efficacité, la rapidité, la sobriété.

Une application d'utilisation fortement ludique laissera au contraire une plus grande place à la fantaisie. La présentation pourra être plus variée, l'objectif étant davantage d'intéresser l'utilisateur par l'outil d'apprentissage que par l'information transmise. Celui-ci doit donc être captivant, attrayant, l'idéal étant de faire passer l'information voulue **sans même que l'utilisateur s'en rende compte**.

- **Axe 2: "initiative laissée à l'utilisateur"**

L'"initiative laissée à l'utilisateur" est un axe indiquant la marge de manoeuvre dont l'utilisateur peut bénéficier. Il constitue donc une indication concernant le **mode de réception de l'information**.

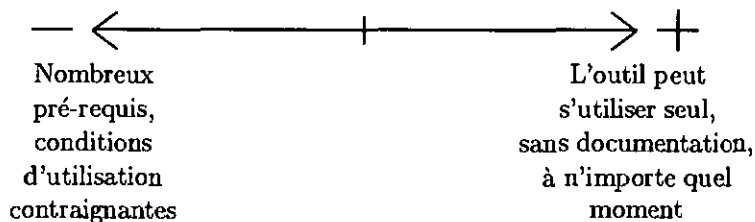


Une application sera fortement directive si le système impose un cheminement rigoureux duquel l'utilisateur ne peut pas, ou très peu fréquemment sortir. Celui-ci se laisse alors guider et adopte une attitude relativement passive puisqu'il n'a pas de liberté de choix.

Au contraire, lorsque la liberté est très grande, l'utilisateur est contraint à effectuer lui-même ses propres choix, à choisir son propre chemin. Il est alors obligé d'amener une réflexion personnelle permettant son orientation. Il est à noter qu'une trop grande liberté peut ne pas être appropriée lorsque l'utilisateur n'a aucune connaissance du domaine à explorer, auquel cas il s'orientera de n'importe quelle façon et se perdra dans un sentiment de confusion complète sans que l'apprentissage puisse lui être bénéfique.

- **Axe 3: "flexibilité de mise en oeuvre"**

La "flexibilité de mise en oeuvre" donne une indication concernant les contraintes liées à l'utilisation de l'outil d'apprentissage.



Une application très flexible est caractérisée par la forte autonomie de l'utilisateur dans son apprentissage. L'outil se suffit à lui-même, il est compréhensible par lui-même. Ces conditions sont indispensables lorsque

l'application est destinée à être utilisée par des personnes de manière strictement individuelle.

Elles sont par contre beaucoup moins importantes lorsque l'application fait par exemple partie d'un cycle d'enseignement où toute l'assistance nécessaire est assurée par un enseignant ou toute autre personne compétente.

Reprenons ces différents axes afin de déterminer quel devrait être le positionnement de notre logiciel dans l'espace des stratégies possibles ainsi définies:

- **Axe 1: "ancrage du réel"**

L'apprentissage par le texte (livre électronique ou dictionnaire électronique) peut être utilisé à des fins purement professionnelles. Dans cette optique, nous préférons la simplicité et la clarté de la présentation à une présentation laissant plus de place à la fantaisie.

Toutefois, l'aspect ludique est un aspect fondamental en vue de faire passer non seulement le **savoir** mais également le **savoir-faire**. Il intervient dans la possibilité d'utiliser les outils statistiques, de modifier des données pour visualiser immédiatement les effets d'un ajout ou d'une suppression de données, de la modification de certaines valeurs sur les résultats fournis par les méthodes statistiques.

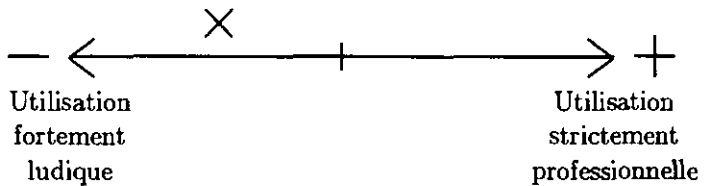
L'utilisation de méthodes graphiques (présentation graphique des résultats à l'écran) permettent également de mieux visualiser les résultats et d'amener un aspect attrayant à l'observation de ces résultats.

Imaginons par exemple l'utilisation des outils statistiques pour effectuer une régression linéaire simple sur deux variables. L'outil permet non seulement d'obtenir les résultats, notamment l'équation de la droite de régression, mais également une visualisation graphique de cette droite.

L'utilisateur, qui a alors la possibilité de modifier la valeur d'une ou plusieurs observations et de constater immédiatement sur le graphe produit à l'écran les effets d'une telle modification, pourra acquérir de manière beaucoup plus rapide et beaucoup plus concrète une compréhension véritable des mécanismes de la méthode de régression, de la validité de cette méthode sur tel ou tel type de données.

Au niveau des possibilités offertes par le logiciel, celui-ci se trouvera donc

positionné davantage du côté d'une utilisation ludique:



• **Axe 2: "initiative laissée à l'utilisateur"**

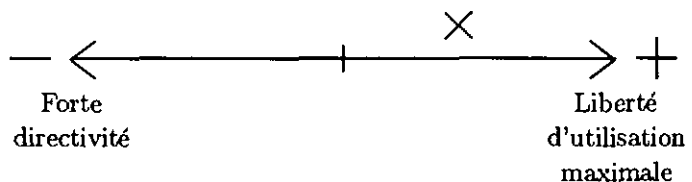
Notre but est de répondre aux besoins et aux désirs de l'utilisateur: un utilisateur débutant aura peut-être besoin d'être guidé, alors qu'un utilisateur plus avancé voudra certainement être beaucoup plus libre dans son apprentissage. Dans cette optique, il est nécessaire d'offrir une grande liberté dans le choix du parcours de la matière.

L'apprentissage par le livre électronique peut se faire:

- par un parcours guidé: sans demande particulière de la part de l'utilisateur, le livre électronique peut être parcouru selon un schéma, une structure prédéterminée.
- par un parcours individuel: une très grande liberté est également laissée à l'initiative de l'utilisateur. S'il le désire, il peut parcourir le livre électronique dans l'ordre qu'il désire, choisir le chemin qui l'intéresse, éviter certains domaines pour passer directement à d'autres domaines qui l'intéressent davantage.

Quant à l'apprentissage par le dictionnaire électronique, il est totalement libre puisqu'il n'impose absolument aucune contrainte à l'utilisateur qui peut voyager absolument comme il le désire au travers de la base de données.

Nous aboutissons donc à un positionnement sur la droite de l'axe:

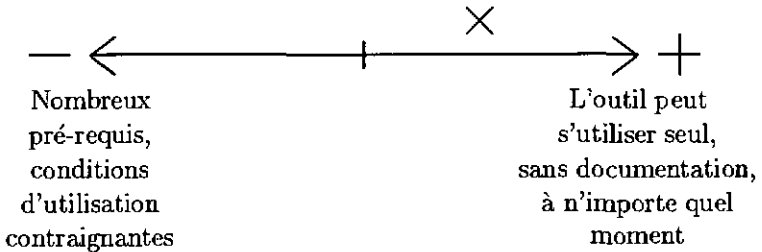


• **Axe 3: "flexibilité de mise en oeuvre"**

L'idéal serait de réaliser un outil d'utilisation totalement indépendante.

Il est clair que le domaine traité nécessite certains pré-requis, notamment dans le domaine des mathématiques. Les notions de mathématique de base, par exemple, ne font pas partie du domaine enseigné. Elles forment un ensemble de connaissances indispensables à l'apprentissage de la statistique. Si ces notions ne sont pas maîtrisées, il sera alors nécessaire de combler ce manque par la lecture d'ouvrages complémentaires.

Toutefois, en dehors de ces pré-requis, l'outil est utilisable de manière totalement autonome, si l'on excepte éventuellement l'utilisation du crayon, de papier et d'une machine à calculer pour effectuer le cas échéant les exercices manuellement avant de les tester sur ordinateur. Ce qui nous amène à positionner le logiciel également sur la droite de cet axe:



L'étude du positionnement du système sur chaque axe individuel nous permet de représenter schématiquement la stratégie pédagogique selon la figure 2.3.

Dans cette représentation synthétique, les carrés (■) représentent les composantes individuelles du logiciel sur chaque axe. Le cercle (●) représente la place du logiciel dans l'espace des stratégies possibles.

Remarquons qu'il n'est pas aisé dans une telle représentation de définir précisément la position du logiciel sur chacun des axes. Une représentation précise ne serait possible que si chaque axe pouvait être composé d'un certain nombre de critères quantitatifs mesurables, ce qui n'est pas le cas au stade de l'élaboration de la stratégie.

L'essentiel est donc de déterminer de quel côté de chaque axe le logiciel doit être positionné, afin de déterminer les principales lignes directrices à suivre

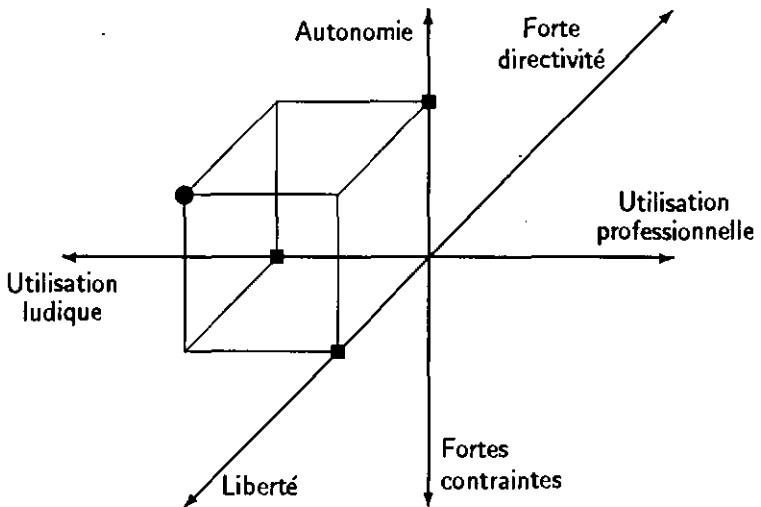


Figure 2.3: Positionnement du système dans l'espace des stratégies possibles

dans la construction du logiciel. En fonction de la figure 2.3, celles-ci peuvent être résumées en trois points:

1. laisser une place importante à l'aspect ludique de l'utilisation du logiciel;
2. encourager l'initiative de l'utilisateur par une liberté d'action maximale;
3. tendre vers un maximum d'autonomie dans l'utilisation du logiciel.

Chapitre 3

Scénario pédagogique

3.1 Introduction

Le scénario pédagogique a pour but de dépasser la formulation des **intentions** pour réaliser une description précise et détaillée des **moyens** choisis pour y satisfaire.

Nous avons jusqu'à présent abouti à une première description globale de notre application, sur la base de l'étude:

- des objectifs;
- du public;
- des sources d'information;
- de la stratégie.

L'étape suivante va nous permettre de préciser la **façon** d'atteindre nos objectifs afin d'aboutir à une description détaillée de tous les éléments nécessaires à la réalisation pratique de l'outil.

Cette étape vise notamment à définir la combinaison des ressources à disposition:

- **ressources en matériel:**
définir la manière de gérer l'écran, structurer la **page-écran** en fonction des possibilités qui s'y rattachent (utilisation de couleurs, de graphiques, etc); cet aspect touche essentiellement à la construction d'une interface homme-machine adaptée d'une part au public visé et d'autre part au média utilisé (l'ordinateur);
- **ressources en contenu:**
structurer l'ensemble des notions à transmettre afin de faciliter l'apprentissage; organiser les opérations permettant l'adaptabilité au cheminement de chaque utilisateur.

3.2 Interface homme-machine

L'objectif de la construction d'une interface homme-machine est d'exploiter au maximum les capacités de l'ordinateur tout en laissant au maximum le champ libre à l'utilisateur. Celui-ci peut alors **communiquer** avec l'ordinateur par un dispositif défini servant d'intermédiaire entre ces deux "systèmes" en présence: l'homme et la machine. La figure 3.1 représente le cycle du flux d'information entre ces deux systèmes.

La définition de l'interface homme-machine est une des phases essentielles du développement d'un système de Formation Assistée par Ordinateur. C'est notamment par l'interface homme-machine que la formation assistée par ordinateur se distingue d'un système de formation classique.

En effet, indépendamment de la qualité des connaissances transmises, une interface bien conçue pourra encourager l'apprentissage, le rendre attractif et instructif, alors qu'une interface mal conçue sera lourde et rébarbative à utiliser.

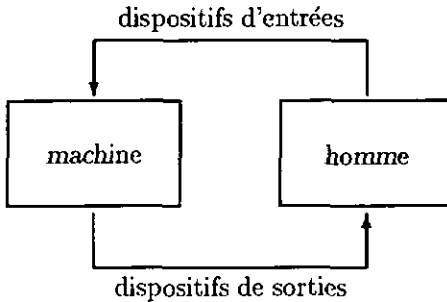


Figure 3.1: Interface homme-machine

La conception d'une interface doit prendre en considération les deux systèmes que l'on désire mettre en relation. Il s'agit donc de considérer le problème d'une part du côté de l'homme et d'autre part du côté de la machine.

1. Interface vue du côté de la machine:

La représentation conceptuelle (abstraite) de l'ensemble des fonctionnalités de l'application se traduit:

- d'une part en une **représentation interne** à la machine (aspects de programmation);
- d'autre part en une **représentation externe** (ce que l'utilisateur reçoit comme information et la manière dont il la reçoit).

Ces deux niveaux de représentation dépendent des contraintes spécifiques du matériel (type d'écran, possibilités d'utilisation d'autres périphériques) et également des contraintes liées aux logiciels ou routines d'entrées/sorties.

2. Interface vue du côté homme:

En vue de rendre une application adaptée à l'utilisateur, il est fondamental de prendre en compte différentes caractéristiques humaines:

- *caractéristiques physiologiques* liées notamment à la perception visuelle;

- caractéristiques de mémorisation liées aux opérations élémentaires réalisées par l'utilisateur;
- système de représentation et de traitement de l'information de l'utilisateur. Ce critère est intimement lié au genre de public à qui l'application est adressée.

La connaissance des caractéristiques humaines est à la base d'une représentation externe adaptée aux besoins et aux capacités de l'utilisateur, à une bonne présentation de l'information.

Le parcours de la littérature en ergonomie sur l'interaction homme-machine permet, selon Barthelet [4], de caractériser l'interface par cinq facteurs principaux:

- **séquence des opérations**
- **dispositifs d'entrées**
- **dispositifs de sorties**
- **temps de réponse**
- **traitement des erreurs**

Chacun de ces éléments peut contribuer au succès ou à l'échec de l'outil en sa qualité de transmetteur de connaissance.

Avant de donner une description plus détaillée de ces facteurs, nous allons fournir quelques considérations ergonomiques qui sont à la base de la conception de l'interface homme-machine.

Le choix d'une définition de l'ergonomie n'est pas aisé car l'accord sur ce point n'est pas général. Nous avons choisi de reprendre la définition de Leplat [28] selon laquelle l'ergonomie est une technologie dont l'objet est l'aménagement des systèmes homme-machine, selon un certain nombre de critères. Ces critères sont par exemple des critères concernant l'opérateur humain:

- la sécurité;
- le confort;
- la satisfaction;

- la réduction de la fatigue.

On dira ainsi qu'un système est ergonomique s'il est bien adapté à son utilisateur du point de vue de ces différents critères.

L'ergonomie des logiciels touche notamment aux problèmes:

- de gestion de l'espace de visualisation, gestion des zones de travail, gestion des couleurs et des effets spéciaux;
- de formatage de l'image;
- de la structure des dialogues;
- de codage de l'information: choix du vocabulaire et des abréviations les plus significatives permettant l'échange homme-machine;
- de gestion du temps;
- de gestion des périphériques, ou moyens de communication, de dialogue (clavier, souris, etc).

Il est important de noter que ce qui est idéal pour un utilisateur débutant ou occasionnel sera souvent désagréable pour un utilisateur averti. Nous pouvons illustrer ce fait par un exemple lié aux moyens de communication entre l'homme et la machine: cette communication se fait fréquemment par l'utilisation des "touches de fonction" du clavier. Toutefois, celles-ci étant en nombre limité (généralement entre 10 ou 12), il n'est pas rare de devoir utiliser des combinaisons de touches, l'utilisation systématique de combinaisons de touches permettant d'effectuer un nombre d'opérations relativement grand. Une telle pratique nécessite une grande mémorisation de la part de l'utilisateur.

Le débutant - de même que l'utilisateur occasionnel - devra sans cesse se référer au manuel d'utilisation pour connaître la fonctionnalité des différentes combinaisons, ce qui engendra une grande perte de temps et éventuellement même un sentiment de réticence vis-à-vis d'un système si peu convivial.

Par contre, l'utilisateur permanent d'un tel système pourra y trouver des avantages et le manipuler avec une grande dextérité.

Il est donc important d'étudier ces problèmes de manière approfondie en vue d'aboutir à des compromis satisfaisants - voire optimaux - dans le plus grand

nombre de situations possibles, même si, comme le soulève Sperandio [51], nous manquons dans ce domaine de modèles universels qui pourraient s'appliquer comme une recette!

3.3 Séquence des opérations

Lors de l'utilisation d'un système de Formation Assistée par Ordinateur, l'étudiant se trouve confronté à une succession d'écrans, chaque écran pouvant correspondre à une opération particulière (lecture de texte, réalisation d'un exercice, etc).

La séquence des opérations est l'enchaînement des opérations tel qu'il est autorisé par l'ordinateur. Sa détermination est un des principaux facteurs caractéristiques de l'interface homme-machine. Elle peut suivre différents modèles plus ou moins directifs, en fonction de la politique choisie de transmission de l'information.

3.3.1 Politiques de transmission de l'information

L'étudiant face à l'écran en situation de Formation Assistée par Ordinateur peut être confronté à deux cas fondamentaux:

1. les données entrées par l'étudiant appellent un commentaire de la machine qui juge de leur validité ou de leur valeur par rapport à une norme. Les données sont donc des "réponses" qui pourront être jugées correctes, erronées, incomplètes, partiellement correctes, etc.

Nous sommes alors dans une situation dite **tutorielle**.

2. les données n'appellent pas de jugement normatif, mais sont traitées par la machine qui exécute à partir d'elles divers calculs selon un modèle préalablement programmé et soumet à l'étudiant le résultat de ses calculs.

C'est l'étudiant qui tire des conclusions en observant les résultats fournis par la machine. Les données sont donc des "ordres" donnés à la machine.

Nous sommes alors dans une situation dite de **simulation**.

Nous allons exposer un peu plus en détail la situation de simulation car c'est celle-ci qui nous intéresse spécifiquement dans le cadre de notre application.

Nous entendons par simulation l'exécution d'un modèle. Une telle situation est caractérisée par le processus suivant:

sollicitation de la machine:
la machine donne un descriptif d'une situation à faire évoluer
et attend de l'étudiant une entrée de données



réponse de l'étudiant:
l'étudiant entre (ou modifie) des données qui ne seront
pas jugées, mais utilisées pour exécuter des calculs



traitement de la machine:
la machine exécute le modèle en calculant les conséquences
des données entrées



rôle de l'étudiant:
l'étudiant observe les résultats pour juger par lui-même
de la valeur ou de la pertinence des données entrées

Dans une situation de simulation, l'utilisateur est donc un "expérimentateur". Il manipule des données et apprend par l'observation et la découverte.

La principale valeur pédagogique d'une telle situation est de promouvoir l'activité de recherche personnelle, et donc de développer l'autonomie de l'étudiant en situation expérimentale qui tâtonne sans maître à disposition pour le guider.

Le choix de la politique de transmission de l'information a une influence directe sur la séquence des opérations proposée par le système. Nous considérons ci-dessous trois types de séquences, allant du plus directif au moins directif.

3.3.2 Chaînage linéaire

Le chaînage linéaire est le modèle pour lequel la directivité est la plus forte. Comme on peut le voir sur la figure 3.2, il n'autorise le passage d'une opération

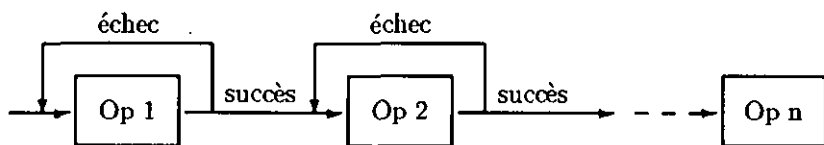


Figure 3.2: Chaînage linéaire

à une autre que lorsque la première a été passée avec succès, quel que soit le sens spécifique donné au terme “succès”:

- passage d'une épreuve ou d'un test;
- résolution d'un exercice;
- lecture d'une information textuelle;
- visualisation d'une démonstration.

Un tel type de séquence des opérations se rencontre notamment dans une situation tutorielle, où l'objectif est d'imposer un parcours fixe d'acquisition d'une certaine connaissance. Il peut s'avérer nécessaire dans le cas de l'apprentissage d'une matière totalement nouvelle pour l'étudiant, lorsque la structure de celle-ci est intimement liée à un ordre naturel d'apprentissage, c'est-à-dire lorsque chaque nouvel élément dépend de l'assimilation de l'élément précédent. Comment, par exemple, enseigner la notion de “soustraction” lorsque celle d’“addition” n'est pas assimilée, ou celle de “division” lorsque la “multiplication” est encore inconnue?

3.3.3 Chaînage individualisé

Le modèle de chaînage individualisé guide également l'étudiant dans son apprentissage. Comme le montre la figure 3.3, la séquence des opérations est fixée selon un réseau bien déterminé. Le parcours de l'étudiant au travers de ce réseau dépend de sa propre compréhension puisqu'un “échec” n'amène pas forcément à la même opération qu'un “succès”.

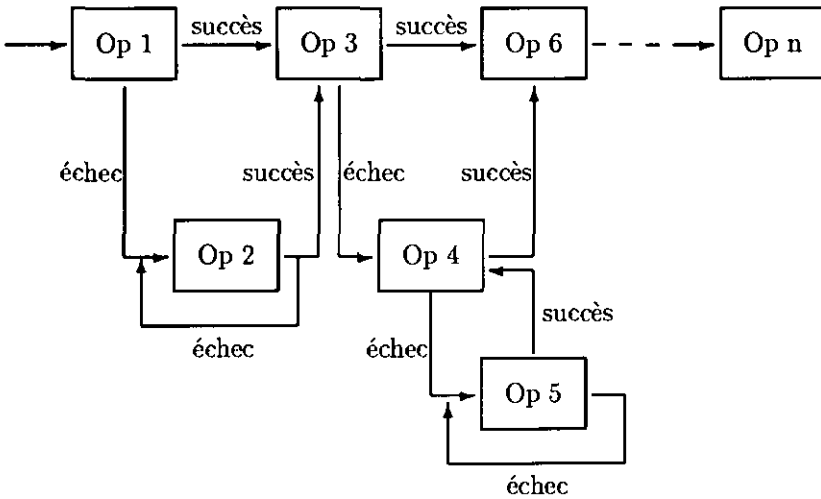


Figure 3.3: Chaînage individualisé

Un tel modèle permet donc d'individualiser l'apprentissage dans le sens où un étudiant plus avancé pourra parcourir plus rapidement l'ensemble du domaine, alors qu'un étudiant moins avancé devra parcourir un chemin plus long lui permettant de mieux approfondir les notions qu'il ne maîtrise pas parfaitement.

Le danger d'un modèle de chaînage individualisé réside dans la complexité de la structure. En effet, une structure trop complexe peut amener une certaine confusion chez l'étudiant qui risque de se perdre et de ne plus percevoir le fil conducteur.

3.3.4 Chaînage modulaire

Dans un modèle à chaînage modulaire, l'étudiant gère lui-même son parcours. Un exemple en est donné en figure 3.4. On peut remarquer que l'étudiant choisit son point d'entrée, et que chaque opération le laisse ensuite libre de ses futurs choix.

C'est une des structures les moins directives et les plus souples. Un tel modèle n'est toutefois adapté que pour certains domaines qui s'y prêtent, et pour un

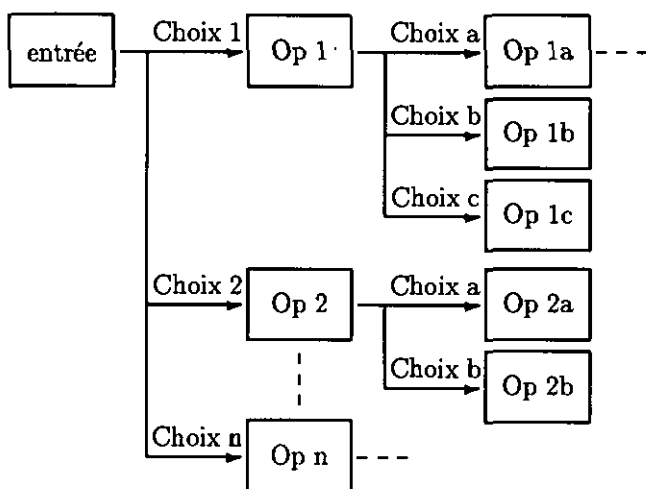


Figure 3.4: Chaînage modulaire

certain type de public (indépendant, capable de discerner ses besoins).

3.3.5 Séquence des opérations relatives à un dictionnaire électronique

D'une manière générale, la consultation d'un dictionnaire consiste à rechercher des mots spécifiques afin d'en lire le contenu. La recherche d'un mot particulier pouvant être nécessaire à la compréhension d'un autre mot, le processus de "retour en arrière" a une place relativement importante.

Le processus global est en fait une suite itérative de choix et de consultation d'un mot, dont la fin ne dépend que de la volonté de l'utilisateur. Si nous considérons ces deux opérations (choisir un mot et consulter un mot) de façon tout-à-fait globale et en faisant totalement abstraction des éléments qui les composent, nous avons donc schématiquement une représentation de la séquence des opérations extrêmement simple. Celle-ci est représentée en figure 3.5.

Dans le dictionnaire électronique, au niveau d'abstraction auquel nous avons

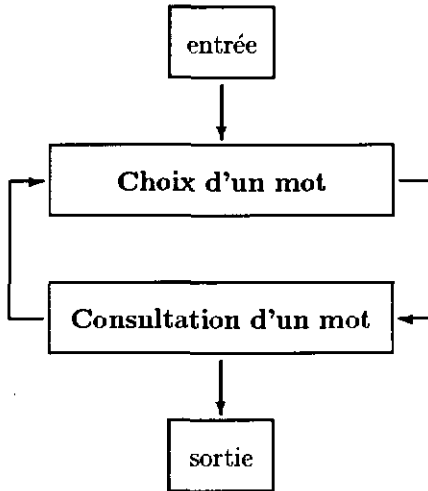


Figure 3.5: Séquence des opérations relatives à un dictionnaire

considéré ces deux opérations fondamentales, le processus est tout-à-fait identique. Toutefois, le contenu de ces deux opérations est fondamentalement différent.

Lors de la consultation d'un dictionnaire, ces deux opérations se résument à une tâche très simple:

- le **choix d'un mot** consiste à rechercher la bonne page dans l'ouvrage;
- la **consultation d'un mot** se résume à la lecture des informations relatives à ce mot.

En revanche, lors de la consultation du dictionnaire électronique, ces deux opérations prennent une toute autre ampleur. Il peut exister de nombreuses possibilités de recherche d'un mot:

- soit par un accès direct au mot, en tapant le mot cherché;
- soit par une sélection du mot dans une liste alphabétique (complète ou partielle) des mots du dictionnaire;

- soit par une sélection dans une liste non alphabétique (une liste par domaine, par centre d'intérêt);
- ou encore par une requête spécifique pouvant amener directement à l'affichage du ou des mots précédemment consultés.

Les possibilités offertes par l'ordinateur sont extrêmement vastes dans ce domaine. Quant à l'opération de consultation d'un mot, celle-ci peut bien évidemment se réduire à l'affichage à l'écran du mot en question. Toutefois, les capacités de l'ordinateur permettent d'aller beaucoup plus loin en offrant par exemple la possibilité d'accéder directement à une information spécifique désirée à l'intérieur d'un mot; ou encore, et c'est là que le terme électronique vient prendre tout son sens, en offrant à l'utilisateur la possibilité de "jouer" avec les concepts présentés. Avec un dictionnaire électronique dans le domaine de la statistique, comme nous l'avons déjà vu, cette notion de "jeu" est tout-à-fait capitale.

3.3.6 Séquence des opérations relatives à un livre électronique

Contrairement au dictionnaire où il n'existe pas de suite logique dans la consultation des différents textes qui le composent, il existe dans un livre un flux logique de parcours de l'ensemble des textes.

Il est possible d'utiliser un livre à la manière d'un dictionnaire (par l'index alphabétique, en ne consultant par exemple que les pages qui concernent une certaine notion) ou même d'en parcourir le contenu de manière totalement désordonnée (consulter les pages 10, 356, 212 puis 478).

Toutefois, nous nous intéressons ici au parcours "normal", ou logique du texte. La structure de base d'un tel parcours est en général séquentielle, mais elle peut également être séquentielle par parties (en l'occurrence par chapitres), dans le sens où la lecture d'un chapitre ne dépend pas forcément ni du chapitre précédent ni du chapitre suivant.

Comme le montre la figure 3.6, on a donc la possibilité d'entrer dans le processus en différents points (en principe les débuts de chapitres), la suite du parcours étant une lecture séquentielle des différentes pages relatives au chapitre.

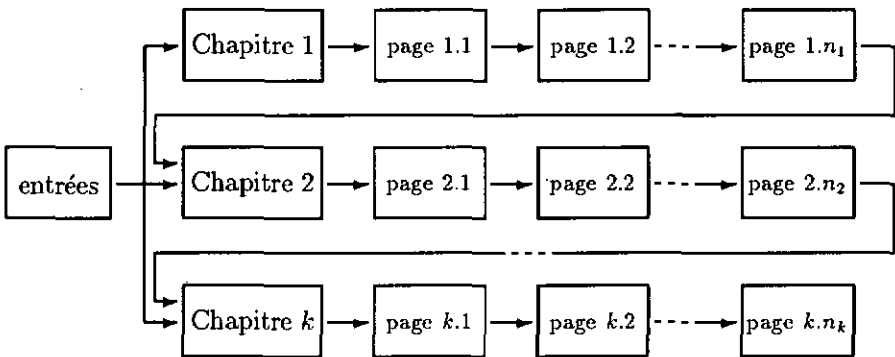


Figure 3.6: Séquence des opérations relatives à un livre

3.4 Dispositifs d'entrées

Les dispositifs d'entrées forment l'ensemble des moyens fournis à l'utilisateur lui permettant d'entrer en contact avec la machine. Les dispositifs d'entrées sont également appelés dispositifs de commandes, la notion de commande étant prise dans un sens général incluant non seulement les ordres donnés par l'utilisateur à la machine, mais également tout autre type d'information (entrée de données ou réponse à une sollicitation de la machine).

L'étude des commandes est également très classique en ergonomie. C'est par les commandes que l'utilisateur fait entrer de l'information dans le système. Elles lui donnent donc le moyen de communiquer avec le système, et le cas échéant de le modifier afin de l'amener à l'état souhaité.

L'étude du dispositif de commandes est liée à deux préoccupations fondamentales: le choix du dispositif (matériel utilisé) et la présentation de l'information donnée à l'utilisateur quant à ses possibilités d'action (ou présentation des commandes).

3.4.1 Choix du dispositif de commandes

Le dispositif de commandes le plus couramment utilisé actuellement est le clavier. Toutefois, la tendance actuelle est d'essayer de concevoir des formes

de dialogue qui diminuent la nécessité de recourir au clavier, les possibilités futures étant davantage tournées vers l'utilisation d'outils tels que la souris, l'écran tactile, le crayon optique, le manche à balais ou la tablette numérisante. Chacun de ces dispositifs a toutefois ses propres champs d'application, ses avantages et ses inconvénients.

Le choix du ou des dispositifs de commandes dépend principalement du type d'application développée. Toutefois, quels que soient les dispositifs choisis, il est bénéfique de prévoir plusieurs possibilités d'entrées, ce qui a l'avantage d'éviter de rester en panne si un des dispositifs est défectueux (ce qui est relativement fréquent en cas d'utilisation du crayon optique par exemple), et également de laisser le choix à l'utilisateur d'employer le dispositif qui lui convient le mieux.

Nous avons pour notre application choisi une solution qui conjugue l'utilisation:

- du **clavier** qui est le dispositif le plus adapté pour toute entrée de réponses ouvertes ou saisie de données;
- de la **souris** qui permet d'effectuer toutes les opérations de choix de commandes, de parcours de listes ou de sélection.

3.4.2 Présentation des commandes

Kowal [26] met en évidence le fait que les commandes disponibles devraient toujours rester visibles à l'écran, essentiellement pour un débutant. Mais il est évident que toutes les commandes ne peuvent pas forcément figurer en permanence à l'écran, surtout si elles sont nombreuses.

Il est alors possible de regrouper les commandes en sous-ensembles dont les noms seuls figurent à l'écran. L'appel à un de ces noms provoquera alors l'affichage dans une partie de l'écran (ou fenêtre) des commandes y relatives.

Cette nouvelle fenêtre peut venir se superposer à d'autres zones de travail qui deviennent alors inactives aussi longtemps qu'un choix n'a pas été effectué. Elle peut aussi correspondre à une zone fixe de l'écran réservée pour ce type particulier d'opération.

Dans l'exemple de la figure 3.7, la zone de commandes occupe en permanence la première ligne de l'écran. Lorsqu'une des commandes est sélectionnée (ici

la commande Bâtit), une fenêtre contenant un ensemble de sous-commandes vient s'ouvrir et se superposer aux autres zones de travail.

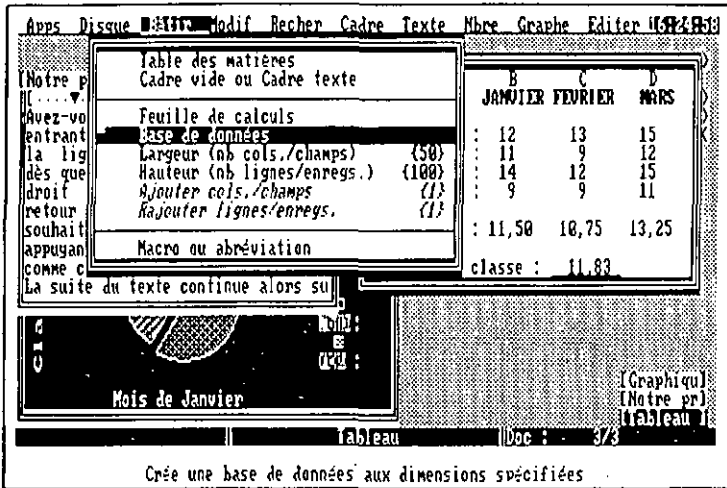


Figure 3.7: Ecran à zones de commandes superposées (Source: *Prise en main de Framework III*, La Commande Electronique, 1988)

Dans la figure 3.8, la zone de commandes est fixe. Elle se situe sur deux lignes en bas de l'écran. Il n'y a pas ici d'empiétement entre les différentes zones. Si la commande sélectionnée appelle également un ensemble de sous-commandes, celles-ci viendront s'afficher dans la même zone qui reste fixe, et recouvrir le premier niveau de commandes.

Le choix du mode de présentation des commandes (en zone fixe ou en superposition) dépend de l'ampleur (nombre de commandes à représenter à l'écran) et de l'organisation de l'ensemble des commandes (possibilité de regroupement des commandes en une structure à deux ou plusieurs niveaux). Si le nombre de commandes est relativement petit et que la structure ne comporte qu'un ou deux niveaux, il est alors possible d'adopter une présentation en zone fixe. On a alors l'avantage d'offrir en permanence l'ensemble des commandes disponibles sur un espace visuel stable.

Toutefois, lorsque le nombre de commandes devient trop important ou que la structure est trop complexe (chaque commande fait obligatoirement appel à tout un embranchement), il devient alors préférable d'opter pour un mode de

présentation en fenêtres superposées.

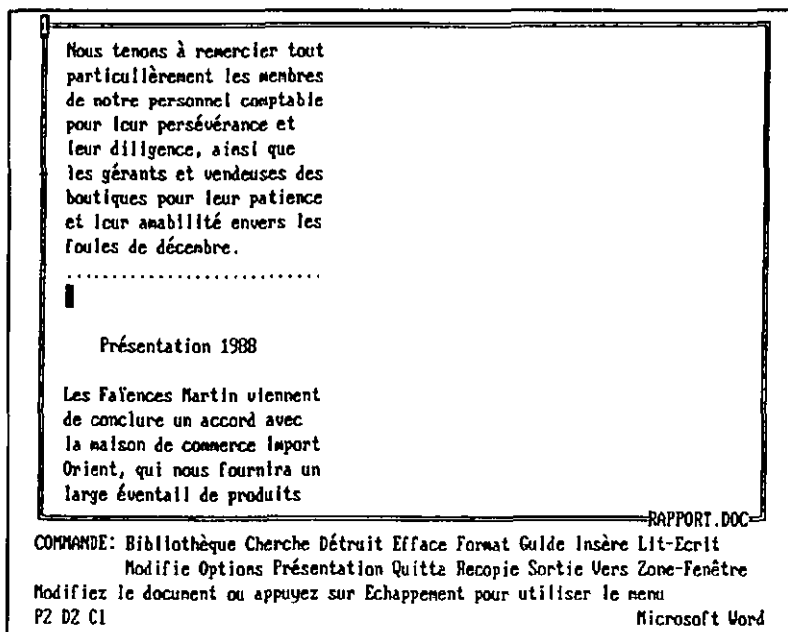


Figure 3.8: Ecran à zone de commandes fixe (Source: *Microsoft Word, guide d'utilisation*, Microsoft Corporation, 1989)

3.5 Dispositifs de sorties

Les dispositifs de sorties sont les éléments de l'interface homme-machine permettant à la machine de communiquer des informations à l'homme.

Nous trouvons notamment l'écran, l'imprimante et les systèmes sonores, ou plus simplement les sons.

Le dispositif de sorties le plus important et le plus utilisé est l'écran. C'est sur celui-ci que nous porterons notre attention.

Nous commencerons par aborder le problème dans sa globalité en proposant certaines règles générales de composition de l'image, l'objectif étant de proposer un écran présentant un aspect équilibré et esthétique. Nous approfondirons

dirons ensuite le problème en étudiant différentes techniques de présentation de l'information, telles que l'utilisation des couleurs, des graphiques et du fenêtrage.

3.5.1 Composition d'un écran

La page-écran a un rôle fondamental à jouer puisque c'est elle qui permet la communication entre le système et l'utilisateur. Il est donc essentiel que celle-ci soit organisée de manière claire et suggestive, qu'elle soit aisément abordable, et qu'elle ne choque pas la sensibilité de l'utilisateur.

Hennart et al. [19] proposent dans leur ouvrage certaines règles de composition d'un écran, desquelles nous retiendrons notamment:

- **le caractère clos des scènes pédagogiques:**

Une scène pédagogique est d'autant mieux assimilée si celle-ci est présentée de manière fixe à l'écran;

- **l'unité du sens de la page:**

Il est important de ne régler qu'un problème par page-écran.

Même si l'explication ne remplit pas un écran, il est préférable d'avoir un écran à moitié vide qu'un écran comprenant plusieurs thèmes, ce qui évite des confusions et augmente la clarté;

- **la clarté du message:**

Tout message transmis doit être clair, dans le sens où il ne doit pas poser de problème de compréhension dû à sa formulation. Dans cette optique, il est nécessaire de veiller à simplifier les formes syntaxiques, et à choisir un vocabulaire simple mais précis;

- **la lisibilité du message:**

Il est important d'aérer la présentation en vue de rendre l'information véhiculée plus lisible. Un texte écrit sera d'autant plus lisible que certaines lois sont respectées:

- l'espace écran doit être stable. Un processus de défilement est beaucoup plus fatigant et pénible pour l'oeil et l'esprit;
- le modèle de présentation doit être permanent. Il est important de définir un écran-type dont la structure reste permanente au travers du logiciel: placer toujours au même endroit et de la même façon

- un titre, une explication, un message d'erreur, etc, facilite le travail de l'utilisateur;
- la présentation de l'information à l'écran doit être aérée en vue de mettre en évidence les informations les plus importantes.

Ces règles ont pour objectif fondamental d'encourager la mémorisation de l'information, en assurant une bonne visibilité, une bonne lisibilité, de même qu'une utilisation facile du système.

L'ensemble de ces considérations nous amène à étudier de manière un peu plus détaillée les différentes techniques de présentation de l'information et la manière d'utiliser ces techniques, le but étant d'aboutir à la composition d'une page-écran qui satisfasse les critères susmentionnés.

3.5.2 Utilisation des couleurs

La couleur est une technique de mise en évidence. Elle constitue un des moyens principaux permettant d'attirer l'attention de l'utilisateur, que ce soit sur une zone particulière de l'écran, sur une phrase spécifique ou sur des mots-clés.

La couleur est un facteur fondamental de perception des signaux, le terme signal étant compris comme représentant l'ensemble des informations que le système fournit à l'utilisateur.

De nombreuses études ont été réalisées à partir des années 60 concernant le rôle joué par les couleurs sur la perception des signaux.

Selon Jones [22], "la couleur semble être un déterminant important de la vitesse de localisation avec une signalisation visuelle dense, même dans les cas où elle est un repère non pertinent". Ceci est une bonne mise en garde contre l'utilisation intempestive de la couleur qui, utilisée à mauvais escient, peut amener à une mauvaise perception de l'information véhiculée.

Cet effet des couleurs a été testé dans de nombreuses études. Smith [49] a démontré qu'une utilisation judicieuse de la couleur permet de réduire le temps de recherche et le temps de comptage d'éléments dans une liste d'environ 75%. Elle permet également de réduire le taux d'erreurs d'un pourcentage équivalent.

Klare et al. [24] ont étudié en 1955 déjà l'effet du **soulignement** en tant que procédé perceptif. Le soulignement permettant la mise en évidence de certains

mots en vue d'attirer l'attention de l'utilisateur sur leur importance joue un rôle perceptif similaire à l'emploi de la couleur.

L'utilisation d'écrans polychromes pose toutefois des problèmes ergonomiques touchant à la psychologie de la perception: quelles couleurs, et combien de couleurs choisir pour coder telles catégories d'information?

Sperandio [51] a fourni quelques réponses à ces problèmes. Il est connu depuis longtemps qu'il faut se limiter à un nombre réduit de couleurs, 9 à 10 selon Sperandio. D'autres auteurs proposent une restriction encore plus grande, un trop grand nombre de couleurs simultanément sur un même écran pouvant amener une sensation de confusion chez l'utilisateur qui ne peut plus discerner d'un seul coup d'œil les informations importantes.

De plus, Sperandio propose certains critères permettant d'effectuer un choix judicieux des couleurs. Voici quelques-uns de ces critères qui montrent bien la complexité de ce choix:

- éviter les couleurs pour lesquelles la perception et la discrimination ne sont pas bonnes (les couleurs situées aux extrémités du spectre visible);
- retenir les couleurs qui "passent bien" à l'écran (et ceci diffère selon les écrans);
- choisir des combinaisons de couleurs qui se différencient bien entre elles;
- respecter les conventions culturelles (le vert, l'orange et le rouge ont souvent des significations d'action qu'il faut respecter).

Tenant compte de ces critères principalement liés à la capacité humaine de perception visuelle, les contraintes physiques de l'écran rendent ce choix encore plus difficile.

Dans le domaine spécifique de la réalisation d'un outil de Formation Assistée par Ordinateur, le rôle des couleurs est double.

1. Rôle didactique:

Les couleurs ont un rôle didactique important à jouer. En effet, l'utilisation de couleurs permet de mettre certaines informations en évidence, d'attirer l'attention du lecteur sur certains points ou certaines informations fondamentales, ce qui favorise leur mémorisation et leur assimilation.

2. Rôle fonctionnel:

Les couleurs utilisées peuvent également avoir un rôle fonctionnel dans le sens où à une certaine couleur peut être rattachée une certaine fonction.

3.5.3 Utilisation des graphiques

Roger et al. [41] ont examiné les effets de la communication graphique dans un dialogue homme-machine. Il en ressort que le graphique est un moyen "naturel" et synthétique de présenter l'information à l'utilisateur. Son rôle est lié à l'image et à son contenu informationnel riche, dense et attractif. Il apporte un élément intuitif non négligeable dans l'enseignement, et peut également constituer un apport motivationnel important.

L'oeil est alors sollicité d'une manière globale, différente de la lecture séquentielle.

Il est toutefois important de remarquer que l'utilisation de graphiques est dépendante des contraintes de matériel, et notamment du type d'écran utilisé.

3.5.4 Utilisation du fenêtrage

L'ergonomie prône la nécessité d'une présentation adéquate de l'information sur l'écran. Le fenêtrage est une technique de présentation de l'information qui peut permettre de clarifier le texte. Il consiste à diviser l'écran en plusieurs zones d'affichage, zones qui peuvent éventuellement se superposer.

Kowal [26] s'est posé des questions quant à l'efficacité du fenêtrage, quant à ses apports positifs sur le plan de la facilitation du dialogue homme-machine. Ces apports se situent notamment sur deux plans:

- au niveau de la méthode de travail de l'utilisateur;
- au niveau de la perception visuelle de l'information.

Selon Miller [34], les gens aiment pouvoir repérer spatialement l'information. Une définition claire de zones de travail au rôle différent permet à l'utilisateur d'adopter une méthode de travail plus rationnelle. Cela lui permet de pouvoir toujours se situer rapidement et efficacement par rapport à l'ensemble des informations qui lui sont fournies à l'écran.

Miller et Thomas [35] préconisent un découpage de l'écran en différents espaces de travail:

- une zone principale de travail;
- une zone de préparation des entrées;
- une zone où sont précisées les opérations du système qui sont en cours;
- une zone de diagnostic et de réponses fixées.

Il est bien évident que le découpage en zones va dépendre de l'application et du type de dialogue y relatif.

Il est à noter que le découpage de l'écran en différentes zones est soumis aux contraintes physiques de l'écran, et notamment à la taille de l'écran. Certaines zones peuvent être fixes, alors que d'autres peuvent se superposer. La figure 3.9 présente un exemple d'écran à fenêtres fixes. Les trois fenêtres figurant sur cet écran ont des emplacements fixes: celle de droite est un espace de saisie de données, la fenêtre la plus importante (en bas à gauche) est un espace de représentation des résultats, et la petite fenêtre longitudinale en haut de l'écran permet de lancer le traitement des données.

Un exemple d'écran à fenêtres superposées est donné en figure 3.10. Dans cet écran, la fenêtre active vient se superposer aux autres fenêtres qui deviennent alors inactives. Ceci permet d'avoir en permanence à l'écran un nombre pratiquement illimité de feuilles de travail.

Bien qu'un écran à fenêtres multiples et superposées apporte certains avantages, aux niveaux notamment de la place gagnée sur l'écran et de l'attractivité d'une telle présentation, l'abus de la superposition de fenêtres devient d'avantage une distraction pour l'utilisateur, voire même une source de confusion.

3.6 Temps de réponse

Le temps de réponse correspond au nombre de secondes écoulées entre le moment où l'utilisateur lance une opération et le moment où l'ordinateur commence à présenter les résultats à l'écran (ou sur l'imprimante). C'est un des critères permettant de juger de l'efficacité d'un logiciel. Il dépend de

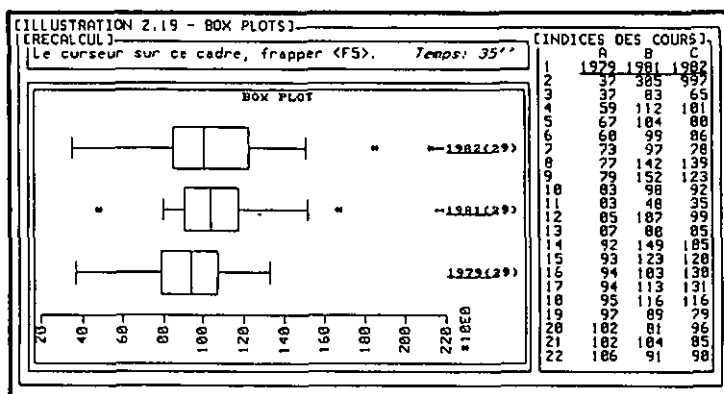


Figure 3.9: Ecran à fenêtres fixes (Source: Ronchetti E., Antille G. et Polla M. *STEP I, Statistique et Probabilités: une Introduction*, Presses Académiques, Neuchâtel, 1989)

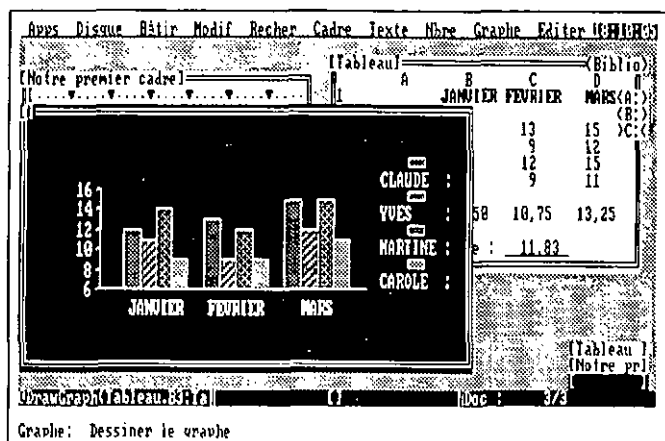


Figure 3.10: Ecran à fenêtres superposées (Source: *Prise en main de Framework III*, La Commande Electronique, 1988)

la structure interne du logiciel (aspects de programmation), mais également du matériel utilisé (type de machine).

Shneiderman [47] propose une discussion détaillée concernant les temps de réponses attendus et acceptés par l'utilisateur dans des contextes très variés. Nous n'en présentons ici que quelques éléments.

Un fait généralement accepté est que le temps de réponse ne doit pas être trop long, un temps de réponse long ayant de multiples implications négatives:

- diminution de la productivité;
- accroissement du taux d'erreurs;
- augmentation de l'insatisfaction et de l'ennui;
- sentiment d'agressivité.

Il n'est toutefois pas très pertinent de vouloir chiffrer le temps d'attente maximal supportable par l'utilisateur, car il dépend de nombreux facteurs.

- **L'expérience passée** a une influence très marquée sur le temps de réponse attendu. Si le temps de réponse correspond à peu près aux attentes de l'utilisateur, il ne manifestera ni satisfaction, ni insatisfaction. Si le temps de réponse est beaucoup plus long, il se produira tout d'abord un sentiment d'inquiétude, puis de l'ennui ou de l'exaspération. Si en revanche le temps de réponse est beaucoup plus court que le temps attendu, l'utilisateur pourra éventuellement mettre en doute les résultats obtenus. Des études ont toutefois montré qu'un temps de réponse plus court augmente généralement le niveau de satisfaction.
- La **tolérance individuelle** joue également un rôle important. Un utilisateur débutant aura d'autres attentes qu'un utilisateur expérimenté, car son temps de réflexion entre deux tâches est généralement plus long.
- Le **type d'opération** demandée est également un facteur important. Si l'on considère par exemple la saisie de texte dans un éditeur, l'affichage à l'écran du caractère tapé au clavier doit être instantané, car le temps de réflexion entre l'appui d'une touche et la suivante est pratiquement nul. Un long temps de réponse dans ce cas particulier rendrait le système inutilisable.

Considérons un tout autre exemple qui appartient au domaine des jeux: si l'utilisateur d'un jeu d'échecs électronique reçoit instantanément la parade de l'adversaire (l'ordinateur) après chacun de ses coups, il sera déstabilisé. Dans un tel cas, le temps de réponse de l'ordinateur est utilisé comme temps de réflexion par l'utilisateur. Une réponse instantanée provoquera certainement une diminution de son temps de réflexion et augmentera les risques d'erreur ou de mauvaise décision.

Cet exemple est un peu particulier, mais il met bien en évidence le fait que le temps de réponse n'est pas forcément qu'un temps d'attente improductif pour l'utilisateur, auquel cas un temps d'attente plus long peut même être mieux perçu qu'un temps d'attente trop court.

Il est intéressant de noter que l'individu se caractérise par une grande capacité d'adaptation, et qu'il peut changer sa manière de travailler en vue de s'accommoder à différents temps de réponse. Si le temps de réponse est long, il va par exemple en profiter pour planifier son travail à venir, ou il va trouver des moyens lui permettant de diminuer le nombre d'interactions avec la machine, lorsque cela est possible.

Toutefois, même s'il s'adapte au temps de réponse qui lui est imposé, un délai dépassant une quinzaine de secondes n'est guère acceptable lorsque l'utilisateur doit rester devant son écran en attendant la réponse.

Pratiquement, il faut admettre que certaines opérations demandent plus de temps d'exécution que d'autres. Ces opérations longues ne doivent pas être supprimées, mais elles doivent pouvoir être détectées afin d'en avertir l'utilisateur. L'ordinateur ne devrait jamais travailler plus de quelques secondes sans afficher un message indiquant à l'utilisateur qu'un travail est en cours. La solution idéale est de lui indiquer le temps approximatif de réalisation de la tâche en cours. Les études concernant ce problème montrent généralement qu'une attente approximative de 5 secondes sans message constitue un seuil d'ennui au-delà duquel s'installe un sentiment de vacuité ou d'inquiétude.

3.7 Traitement des erreurs

Le problème de traitement des erreurs concerne la facilité qu'a l'utilisateur pour reconnaître et rectifier ses erreurs, quelle qu'en soit la nature.

Le rôle des messages d'erreur n'est pas uniquement d'indiquer à l'utilisateur

qu'il a commis une faute. Ceux-ci doivent également lui indiquer quel type d'erreur il a commis et même éventuellement pourquoi il a commis cette erreur, et lui donner des indications concernant les rectifications à effectuer.

Des messages d'erreur du genre **ERREUR DE SYNTAXE** ou **ENTREE NON VALIDE** sont beaucoup trop génériques. Ils indiquent qu'une faute a été commise, mais ne fournissent aucun élément permettant à l'utilisateur de connaître la cause exacte de son erreur et de la rectifier. Il est préférable dans un tel cas d'offrir un message indiquant quelle est la syntaxe correcte, ou quelles entrées sont permises.

Certains systèmes offrent uniquement un code d'erreur faisant référence à un numéro d'erreur expliquée dans le manuel d'utilisation. Cette solution est souvent très désagréable, car l'utilisateur n'a pas forcément le manuel sous la main, et de plus, cela nécessite un temps de recherche qui peut déranger dans la continuité du travail. Ce genre de message est à éviter, car il n'offre que très peu d'aide, surtout à l'utilisateur débutant.

Toutefois, si une trop grande sobriété dans les messages n'est pas recommandable, des messages trop longs ou trop fantaisistes ne constituent pas non plus une aide pour l'utilisateur. De tels messages sont parfois bien acceptés par l'utilisateur débutant qui leur trouve un certain attrait, mais ils deviennent très vite agaçants pour l'utilisateur un peu plus expérimenté, et empêchent la fluidité du travail.

Si le contenu du message est important, le ton sur lequel celui-ci est rédigé est également capital. Il est inutile de vouloir "condamner" l'utilisateur en lui infligeant des messages rudes ou désobligeants. Un ton pénalisant, plutôt que d'encourager l'utilisateur à persévérer, aura davantage tendance à le frustrer ou à le placer en situation de stress.

Certaines applications émettent un signal sonore à chaque apparition d'une erreur. Ce procédé peut être utile si l'utilisateur aurait pu ne pas remarquer son erreur, toutefois, il peut être également très embarrassant si d'autres personnes travaillent dans la même salle. Il est en général préférable de laisser à l'utilisateur le choix de travailler avec ou sans signal sonore.

En résumé, voici différents critères proposés par Shneiderman [47] et qui devraient servir de guide dans la construction des messages d'erreurs:

- être le plus précis possible;

- être constructif, en indiquant quels sont les moyens de corriger l'erreur;
- utiliser un ton positif;
- choisir des phrases centrées sur l'utilisateur;
- considérer éventuellement différents niveaux de message (message plus complet sur demande de l'utilisateur);
- garder une constance dans les formes grammaticales, la terminologie et les abréviations;
- garder une constance dans le placement des messages et leur présentation.

3.8 Structuration du contenu

Nous avons donné en section 2.6 une première description très globale du domaine statistique que nous voulons transmettre au travers d'un outil de Formation Assistée par Ordinateur. Nous allons dans cette section étudier ce domaine de manière un peu plus approfondie, en déterminant d'une part quelles notions spécifiques nous voulons transmettre, et d'autre part quelles sont les relations entre ces notions.

L'étude des relations est fondamentale dans le but d'aboutir à une bonne perception de l'information par l'utilisateur. Dans cette optique, il est essentiel d'avoir non seulement une bonne connaissance théorique du domaine, mais également une bonne connaissance normative.

Ceci implique de connaître non seulement la structure globale du contenu, mais également l'imbrication de ses différents éléments (ou des différentes notions qui le composent).

Nous pouvons distinguer deux types de relations entre les notions:

- les relations internes, ou liens de dépendance;
- les relations externes, ou mots-clés.

L'utilité de la détermination des liens de dépendance est de fournir le flux d'apprentissage dans un système d'apprentissage structuré (livre électronique),

et d'offrir la possibilité de situer une notion particulière dans son contexte dans un système d'apprentissage ponctuel (dictionnaire électronique).

Les mots-clés peuvent toujours être donnés comme références, que ce soit au cours d'un processus d'apprentissage structuré ou lors de la consultation ponctuelle d'une notion particulière.

3.8.1 Relations internes

L'étude des relations internes du contenu consiste à ordonner les notions selon leurs relations de dépendance. (La notion **B** dépend de la notion **A** si l'apprentissage de la notion **B** nécessite la connaissance préalable de la notion **A**).

La structure des relations ainsi définie va servir de guide à l'enseignement de l'ensemble du domaine traité.

Les structures possibles de représentation des relations internes sont innombrables. Toutefois, les plus typiques sont la structure séquentielle, la structure hiérarchique et la structure en réseau. Voici une brève description de ces trois types de structures:

1. Structure séquentielle

Comme le montre la figure 3.11, la structure séquentielle est un cas extrêmement simple. Dans une telle structure de dépendance, il est nécessaire d'acquérir le premier élément, puis le second, etc. Nous parlons ici du concept d'élément dans un sens tout-à-fait général, un élément pouvant faire référence à une notion bien précise, ou au contraire à tout un domaine impliquant la compréhension de nombreuses notions individuelles.

Une structure peut donc être représentée globalement sous une forme séquentielle alors que chaque élément de cette séquence peut avoir sa propre structure (qui peut ne pas être séquentielle).

Une telle structure est adaptée à un domaine d'enseignement lorsque chaque élément appartenant à ce domaine constitue un pré-requis à l'assimilation de l'élément suivant.

Si l'on recouvre un domaine suffisamment restreint, pratiquement chaque domaine peut être représenté par ce type de structure.

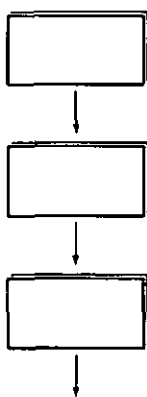


Figure 3.11: Structure séquentielle

2. Structure hiérarchique

Une structure hiérarchique est représentée par une arborescence, comme l'illustre la figure 3.12.

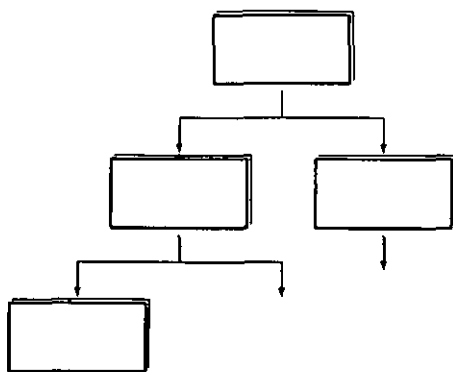


Figure 3.12: Structure hiérarchique

Dans une telle structure, l'apprentissage d'un élément peut mener à l'apprentissage d'un ou de plusieurs autres éléments indépendants. A chaque étape doit donc s'effectuer un choix.

Cette structure correspond à un domaine où certaines notions sont à acquérir à la base, puis différents embranchements permettent d'aborder

différentes autres branches indépendantes (mais formant un tout).

On pourrait également imaginer une structure hiérarchique inversée dans le cas où l'objectif final de l'apprentissage est l'acquisition d'une notion particulière. On aurait une telle structure si cette notion finale dépend de plusieurs autres notions indépendantes entre elles, elles-mêmes dépendant d'une ou de plusieurs autres notions indépendantes entre elles, et ainsi de suite.

Il est à remarquer que la structure linéaire est un cas particulier de la structure hiérarchique (lorsqu'à chaque niveau, l'apprentissage d'un élément ne peut mener qu'à l'apprentissage d'un seul autre élément).

3. Structure en réseau

Le cas du réseau est une des représentations les plus complexes. Un exemple en est donné par la figure 3.13.

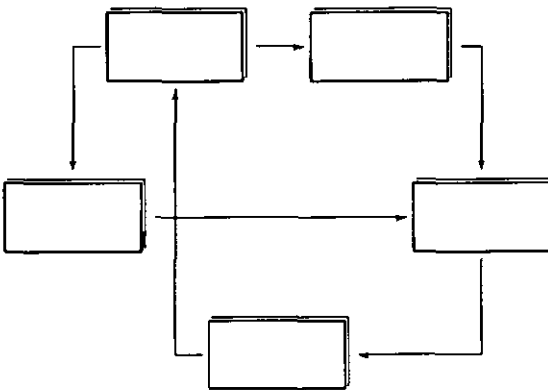


Figure 3.13: Structure en réseau

Dans une telle structure, de nombreux éléments sont en interdépendance, ce qui rend sa compréhension et son assimilation complexes.

Il est très souvent possible de présenter un même domaine selon plusieurs structures différentes. Lors de la construction de la structure des relations internes, il faudra tenter de trouver un compromis entre:

- une structure très complète, représentant absolument toutes les relations internes, mais dont la complexité ne permet plus d'avoir une compréhension globale du système;
- et une structure très simple, mais beaucoup trop schématique car la subtilité des relations n'y apparaît plus du tout.

La structure ainsi établie ne sera utile dans le processus d'enseignement que si elle est d'une part suffisamment simple pour que l'on puisse facilement s'y positionner et savoir où l'on se trouve, et d'autre part si elle est suffisamment complète pour fournir une image instructive et approfondie du domaine.

3.8.2 Construction des relations internes

Nous allons présenter dans cette section un exemple de processus de construction des relations internes. Nous nous baserons sur un domaine restreint, qui inclut quelques notions de base d'apprentissage de la statistique descriptive.

Le processus de détermination de la structure des relations consiste tout d'abord à établir la liste des notions à acquérir (ou à transmettre), puis à effectuer un regroupement des notions en différentes familles. La dernière étape vise à étudier de manière détaillée les relations entre les notions à l'intérieur de chaque famille.

1. Etablissement de la liste des notions à acquérir

Dans le tableau 3.1, nous présentons une liste des notions statistiques à acquérir, avec pour chaque notion la ou les relations directes de dépendance vis-à-vis d'autres notions. Ces relations représentent donc les connaissances indispensables à l'assimilation de la notion.

Les notions listées dans le tableau correspondent à l'ensemble des notions qui sont traitées et utilisées dans une partie du domaine de la statistique descriptive. La liste n'est pas fournie par ordre alphabétique, mais est déjà ordonnée de telle sorte qu'elle fasse apparaître le mieux possible la structure des différentes relations.

Les relations indiquent pour chaque notion quels sont leurs principaux liens de dépendance par rapport aux autres notions. Si nous prenons par exemple le mot *coefficient de variation* (22), nous trouvons dans la colonne des relations que celui-ci est lié à trois autres notions: *mesure de*

	Notion	Relations
1	Mesure de location	
2	Mesure de tendance centrale	1
3	Mesure de dispersion	2
4	Mesure de forme	2,3
5	Mesure d'aplatissement	2,3,4
6	Mesure d'asymétrie	2,3,4
7	Moyenne	1,2
8	Moyenne arithmétique	1,2,7
9	Moyenne arithmétique pondérée	1,2,7,8
10	Moyenne géométrique	1,2,7
11	Moyenne harmonique	1,2,7
12	Quantile	1
13	Quartile	1,12
14	Décile	1,12
15	Centile	1,12
16	Médiane	1,2,12
17	Mode	1,2
18	Déviation	3
19	Ecart moyen	3,8,18
20	Ecart-type	3,8,18
21	Variance	3,8,18
22	Coefficient de variation	3,8,20
23	Empan	3
24	Intervalle interquartile	3,13
25	Coefficient β_1 de Pearson	4,6,20
26	Coefficient β_2 de Pearson	4,5,20
27	Coefficient de Yule et Kendall	4,6,13,16

Table 3.1: Liste des notions

dispersion (3), *écart-type* (20) et *moyenne arithmétique* (8). Cela signifie que ces trois notions doivent être connues avant de pouvoir procéder à l'apprentissage du *coefficient de variation*.

2. Regroupement des notions en familles

L'étude des relations du tableau précédent doit nous permettre de mettre un peu d'ordre dans les différentes notions afin de dégager des groupes de notions appartenant à une même "filère". Nous appelons un tel groupe une **famille**.

Nous avons déterminé un regroupement en cinq familles. Celles-ci sont présentées dans le tableau 3.2.

1	Mesures de tendance centrale	notions 1, 2, 7 à 11, 16, 17
2	Quantiles	notions 1, 12 à 15
3	Mesures de dispersion	notions 3, 18 à 24
4	Mesures d'aplatissement	notions 4, 5, 26
5	Mesures d'asymétrie	notions 4, 6, 25, 27

Table 3.2: Regroupement des notions en familles

En fonction de cette classification, les familles ne sont pas strictement indépendantes les unes des autres.

Nous pouvons constater que certaines notions d'une famille dépendent de notions d'autres familles; par exemple *écart-type* (20), classé dans la famille 3, dépend de *moyenne arithmétique* (8), classée dans la famille 1.

Par ce processus de regroupement en familles, nous perdons donc un peu d'information (car certaines relations sont perdues dans le regroupement), mais nous gagnons en clarté, car la structure de chaque famille peut être représentée de manière simple, et expressive.

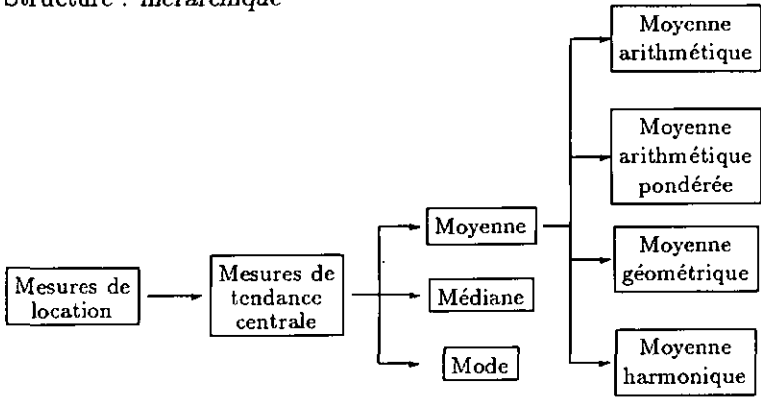
3. Etude des relations à l'intérieur des familles

Sur la base du tableau des notions et de leurs relations, cette étape consiste à représenter l'organisation des différentes notions à l'intérieur de chaque famille.

En fonction des relations décrites pour notre exemple dans le tableau 3.1, nous proposons pour chaque famille la structure suivante:

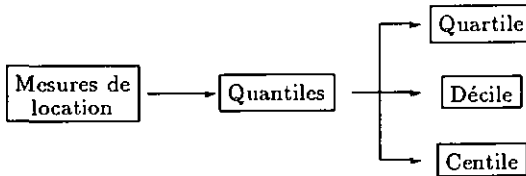
• **Famille 1: Mesures de tendance centrale**

Structure : *hiérarchique*



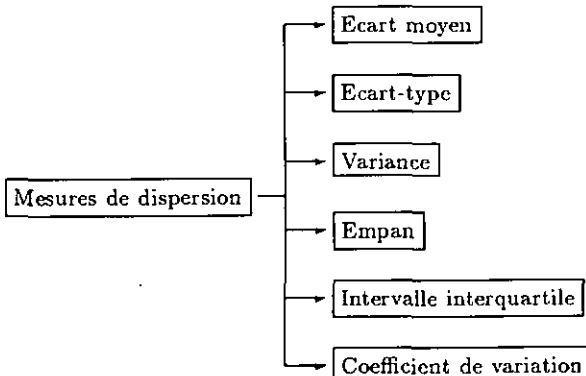
• **Famille 2: Quantiles**

Structure : *hiérarchique*



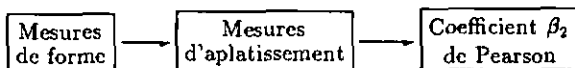
• **Famille 3: Mesures de dispersion**

Structure : *hiérarchique*



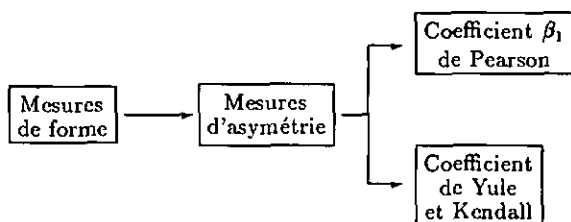
- **Famille 4: Mesures d'aplatissement**

Structure : *linéaire* (cas particulier de la structure hiérarchique)



- **Famille 5: Mesures d'asymétrie**

Structure : *hiérarchique*



Il n'est pas exclu dans cette étape de simplifier les relations dans une optique d'homogénéité, le but étant d'aboutir à une structure du même type pour toutes les familles. Ceci constitue non seulement un apport de clarté pour l'utilisateur, mais également un avantage pour le développeur qui pourra ainsi gérer chaque famille de manière identique.

Nous pouvons constater dans notre exemple que les cinq familles sont caractérisées par une structure **hiérarchique**, c'est-à-dire qu'elles ont un ancrage dont dépendent toutes les autres notions.

3.8.3 Relations externes

En plus des relations internes (ou liens de dépendance), les relations externes nous aident à mieux comprendre comment les différents éléments d'un ensemble sont reliés entre eux.

Les relations externes relatives à un mot (ou mot de base) fournissent l'ensemble de ses **mot-clés**, un mot-clé étant une notion qui peut être liée au mot de

base en fonction de plusieurs critères:

Critère de relation	Exemple
le mot-clé se rapproche du mot de base par son sens (sa signification)	erreur et résidu
le mot-clé, bien qu'il ait un sens totalement différent du mot de base, s'utilise dans les mêmes domaines, dans les mêmes contextes	coefficient de corrélation et régression linéaire s'utilisent tous deux dans le contexte de l'analyse de régression
le mot-clé est un "sous-ensemble" du mot de base	le mot loi de probabilité continue est un des aspects possibles du mot loi de probabilité
le mot-clé "contient" le mot de base	le mot quantile est une famille de méthodes qui contient celle de quartile
le mot-clé aide à la compréhension du mot de base	la notion d'axe aide à la compréhension de la notion de graphe
le mot-clé est une alternative au mot de base	la médiane peut être utilisée au même titre que la moyenne arithmétique comme mesure de tendance centrale

En fonction de ces critères, une notion est toujours liée à plusieurs autres, et parfois même selon des relations multiples, au sein de réseaux complexes. Si l'on considère par exemple le mot *moyenne*, c'est une notion qui apparaît dans la quasi totalité des théories et des méthodes statistiques. Si nous appliquions les critères définis ci-dessus de manière très rigoureuse, l'ensemble des mots-clés relatifs au mot *moyenne* serait donc extrêmement vaste. Il convient alors de faire un tri parmi l'ensemble des notions qui pourraient répondre à ces différents critères et de n'en retenir que ceux qui paraissent les plus pertinents.

Chapitre 4

Description de l'application

4.1 Introduction

Dans les chapitres précédents, nous avons exposé les différents problèmes qui se posent de manière générale dans l'élaboration d'un outil de Formation Assistée par Ordinateur, tout en approfondissant les aspects qui touchent plus particulièrement le domaine d'application qui nous intéresse.

L'ensemble des points qui ont été étudiés dans le chapitre précédent nous ont permis de définir les moyens dont nous disposons en vue de réaliser un outil qui conjugue les qualités de l'utilité pédagogique et de la solidité scientifique des connaissances véhiculées.

Nous sommes dès lors en mesure de présenter une description de notre application de façon plus détaillée, en mettant en évidence nos objectifs et les

différents moyens que nous avons mis en oeuvre en vue d'atteindre ces objectifs.

Comme nous l'avons déjà mentionné en section 2.7, l'application globale est composée de trois modules: un livre électronique et un dictionnaire électronique, tous deux pouvant faire appel au troisième module: les outils statistiques.

Chacun des deux premiers modules, associé au troisième, pourrait être considéré comme un outil autonome. Toutefois, l'accès au dictionnaire électronique par le livre électronique peut être d'une grande utilité, raison pour laquelle nous préférons offrir les deux modules en un seul logiciel permettant ainsi un passage aisé entre l'un ou l'autre des modules.

L'ensemble du logiciel a été développé dans une optique d'indépendance d'utilisation. Nous avons donc tourné nos efforts vers le développement d'un outil qui puisse s'utiliser de manière totalement autonome, ne nécessitant pas d'apport extérieur. Toutefois, l'utilisation du logiciel exige naturellement certains pré-requis. Ceux-ci concernent d'une part le contenu (aspects mathématiques et statistiques), et d'autre part le maniement de l'outil. Les connaissances préalables de l'utilisateur doivent donc pouvoir répondre à:

- **des pré-requis informatiques:**
ceux-ci se réduisent à la capacité de communiquer avec l'ordinateur par l'intermédiaire d'un clavier et d'un écran, de même qu'au respect des consignes d'appel du programme principal, la suite des opérations étant dirigée par des messages indiquant à l'utilisateur les différentes fonctions à sa disposition;
- **des pré-requis statistiques:**
des connaissances de base sont nécessaires lors de l'utilisation du dictionnaire électronique (l'utilisateur doit être capable de déterminer quels sont ses besoins afin d'accéder à l'information voulue);
- **des pré-requis mathématiques:**
l'apprentissage de la statistique, au niveau où nous le proposons, nécessite des connaissances de mathématiques de base.

En dehors de ces exigences préalables, l'ensemble de l'application que nous proposons a été développé dans un constant souci d'aide à une utilisation totalement autonome et individuelle.

La section suivante expose quelques aspects généraux concernant le mode de

présentation de l'information adopté, valable conjointement pour les trois modules. Une description plus détaillée est ensuite donnée pour chaque module séparément.

4.2 Présentation de l'information

Les trois modules de l'application globale ont été développés sur la base des mêmes règles de définition des dispositifs d'entrées et de sorties d'information, en vue d'assurer une homogénéité de présentation et de dialogue.

Comme le montre la figure 4.1, nous proposons l'utilisation d'une page-écran standard divisée en différentes zones, à chaque zone étant rattachée une fonctionnalité particulière:

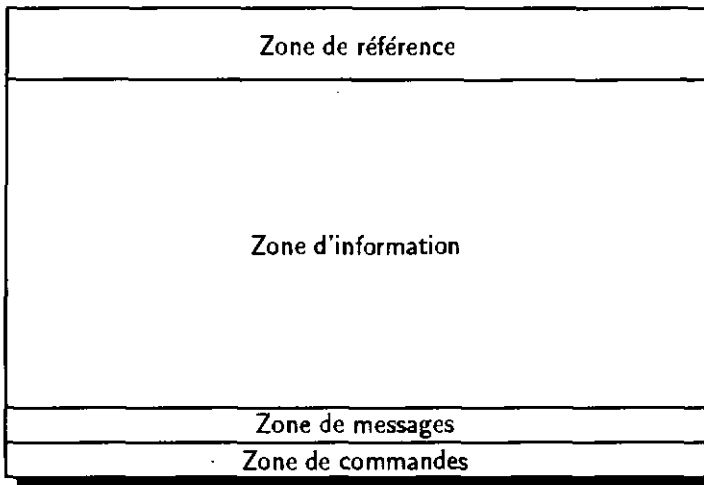


Figure 4.1: Division de l'écran en zones

• Zone de référence

Cette zone permet à l'utilisateur de connaître en permanence sa position dans le logiciel. La première indication fournie est donc naturellement le module actif (le dictionnaire, le livre ou les outils statistiques). En fonction du module, la position exacte est donnée par:

- le mot consulté, ainsi que sa traduction anglaise, de même que la

rubrique en cours dans le dictionnaire;

- le chapitre, la section ou sous-section, ainsi que le numéro de page dans le livre;
- le nom de l'outil dans les outils statistiques.

• Zone d'information

La zone d'information constitue la plus grande surface de l'écran. Elle est utilisée:

- comme zone d'affichage du texte relatif à une rubrique lors de la consultation d'un mot du dictionnaire;
- comme zone d'affichage du texte relatif à un chapitre lors de la consultation du livre;
- comme zone de saisie des données et d'affichage des résultats lors de la réalisation d'exemples personnalisés par l'utilisation des outils statistiques.

• Zone de messages

Cette zone est strictement réservée au système pour l'affichage des messages fournis à l'utilisateur. C'est une zone fixe qui occupe l'avant-dernière ligne de l'écran. Ces messages peuvent être des messages d'erreur ou des informations concernant la marche à suivre à un instant donné.

• Zone de commandes

La zone de commandes occupe en permanence la dernière ligne de la page-écran standard. Elle donne continuellement à l'utilisateur une indication concernant les commandes disponibles à un moment donné en vue de communiquer avec le système. Elle fournit la liste des différentes touches de fonction actives suivies d'un bref label indicatif de leur fonctionnalité.

L'information susceptible d'être affichée à l'écran est essentiellement du texte, que ce soit dans le dictionnaire ou dans le livre électronique. Nous avons donc choisi de concevoir la page-écran standard en mode texte. Ce choix a certains inconvénients, le plus important d'entre eux étant la difficulté de représentation des indices et des exposants dans la rédaction des formules mathématiques. Toutefois, le mode texte possède également des avantages non négligeables. Il permet notamment un affichage plus rapide et une gestion beaucoup plus aisée de l'affichage "déroulant" des textes (scrolling). De plus,

les caractères y sont de meilleure définition qu'en mode graphique, ce qui facilite la lecture à l'utilisateur.

Nous utilisons néanmoins également le mode graphique en vue d'illustrer certaines explications par des graphiques ou des schémas, de même que dans les outils statistiques aboutissant à une représentation graphique des résultats.

4.3 Outils statistiques

4.3.1 Objectifs des outils statistiques

Les outils statistiques forment l'ensemble des moyens que nous avons mis en oeuvre afin d'offrir une possibilité d'apprentissage pratique, actif et dynamique de la statistique. Leur but est donc de faire acquérir un certain **savoir-faire** dans l'utilisation des méthodes qui sont exposées sous forme de texte dans le dictionnaire et le livre électroniques.

Ils offrent la possibilité de manipuler des données, de jouer avec les méthodes afin d'apprendre et de comprendre par l'essai, par l'observation, par la découverte et l'expérimentation, et non seulement par la lecture passive de textes.

Dans la plupart des logiciels statistiques classiques, les procédures de calcul sont totalement opaques pour l'utilisateur: il entre des données, appelle une procédure de traitement de ses données, et reçoit les résultats du traitement demandé.

Notre objectif n'est pas de recréer un tel logiciel statistique. Les outils statistiques que nous proposons ont été développés dans une optique d'aide à l'apprentissage.

Cela signifie que le but premier d'un tel outil n'est pas de fournir des résultats à l'utilisateur, mais plutôt de lui montrer comment ces résultats ont été générés, c'est-à-dire:

- quelle formule de calcul a été employée pour aboutir au résultat;
- quels calculs intermédiaires ont dû être effectués, et quels sont les résultats de ces calculs.

Dans un but pédagogique d'attractivité et afin d'apporter une aide à une

meilleure compréhension, nous avons également, dans la mesure du possible, illustré les résultats par des images graphiques.

Très globalement, les outils statistiques permettent à l'utilisateur de:

- saisir des données sous différentes formes (données brutes, groupées par valeur ou groupées par classes) selon les outils;
- corriger ses données (en ajouter, en supprimer ou les modifier);
- sauvegarder ses données dans un fichier et relire les données d'un fichier préalablement sauvegardé;
- traiter ses données et en obtenir les résultats numériques;
- consulter la ou les formules utilisées lors du traitement;
- obtenir des explications concernant le déroulement du traitement (calculs intermédiaires, messages explicatifs en cas d'erreur (impossibilité) de traitement);
- obtenir des explications graphiques (lorsque cela peut aider à la compréhension);
- imprimer ses données ainsi que l'ensemble des résultats fournis par l'ordinateur (résultat du traitement, formules, explications, graphiques).

4.3.2 Dispositifs de dialogue

L'outil principal de communication entre l'utilisateur et la machine est la zone d'information de la page-écran standard. Cette zone est divisée en deux parties dans les outils statistiques:

1. la moitié supérieure de la zone est réservée à toutes les opérations de saisie des données. Les procédures de saisie permettent d'entrer un maximum de 100 lignes de données dans un cadre déroulant (technique du "scrolling"). La taille du cadre est fixe et permet de visualiser simultanément 7 lignes de données. La forme du cadre dépend du type de données dont voici quelques exemples:

(a) saisie de données brutes:

i	Valeurs x_i
1	
2	
3	
4	
5	
6	
7	

(b) saisie de données groupées par classe:

i	Borne inf.	Borne sup.	Fréq. f_i
1			
2			
3			
4			
5			
6			
7			

(c) données relatives à une variable dépendante Y et une variable indépendante X :

i	Y (var. dép.)	X (var. indép.)
1		
2		
3		
4		
5		
6		
7		

2. la moitié inférieure est une zone d'affichage de l'information (résultats, formules, explications). Lorsque l'étendue de cette zone est insuffisante pour l'affichage de l'ensemble des résultats, des écrans supplémentaires complets sont également utilisés.

Considérons un exemple donné par l'outil permettant d'effectuer une analyse de régression linéaire simple.

La figure 4.2 correspond à la page-écran standard. La zone de référence (première ligne) indique le nom de l'outil utilisé, la zone de commandes (dernière ligne) présente les différentes fonctionnalités offertes par l'outil, et la zone de messages (avant-dernière ligne) fourni les éventuels messages à l'utilisateur.

Outil: REGRESSION LINEAIRE SIMPLE		
i	Y (var. dép.)	X (var. indép.)
1	42.000	14.000
2	33.000	10.000
3	38.000	13.000
4	28.000	7.500
5	34.000	11.100
6	50.000	19.000
7	42.000	13.000

RESULTAT:	FORMULES:
Droite estimée:	$Y^* = a^* + b^* \cdot X$
$Y^* = 13.385 + 1.991X$	$a^* = \bar{Y} - b^* \cdot \bar{X}$
	$b^* = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$
	$n = \text{nombre d'observations}$

Tapez <RETURN> pour consulter le graphique, ou <ESC> sinon

F1: DONNEES F2: TRAITEMENT F3: FORMULE F4: TESTS F5: IMPRESSION ESC

Figure 4.2: Page-écran standard d'un outil statistique

Les données sont saisies interactivement dans le cadre prévu à cet effet dans la moitié supérieure de la zone d'information, et les premiers résultats de l'analyse sont présentés dans la moitié inférieure de cette zone.

Sur demande, l'utilisateur a la possibilité d'accéder à d'autres écrans apportant des informations complémentaires. Certains écrans peuvent être en mode texte, d'autres en mode graphique.

Dans notre exemple, des informations complémentaires sont données par le tableau d'analyse de variance présenté en figure 4.3, et par le tableau des tests sur les coefficients de régression présenté en figure 4.4.

Un écran complémentaire en mode graphique est donné par la figure 4.5 qui fournit une représentation graphique des observations et de la droite de régression estimée.

Outil: REGRESSION LINEAIRE SIMPLE				
Tableau d'analyse de variance				
Source de variation	Somme des carrés	d.l.	Moyenne des carrés	F
Régression	3.338503E+02	1	3.338503E+02	
Résidua	4.264970E+01	8	5.331213E+00	62.622
Total	3.765000E+02	9		

REGLE DE DECISION:

Rejeter H_0 si $F(\text{calculé}) > F(\text{table})$
 Accepter H_0 si $F(\text{calculé}) \leq F(\text{table})$
 $F(\text{table}) = 5.32$
 -> H_0 rejetée

Tapez <RETURN> pour consulter les tests sur les coefficients, ou <ESC> sinon

Figure 4.3: Outils statistiques: écran texte complémentaire (1)

Outil: REGRESSION LINEAIRE SIMPLE				
	Coefficient	Ecart-type	t	Valeur P
Constante	13.385	3.134	4.271	0.003
X	1.991	0.252	7.913	0.000

Tapez <RETURN> pour continuer

Figure 4.4: Outils statistiques: écran texte complémentaire (2)

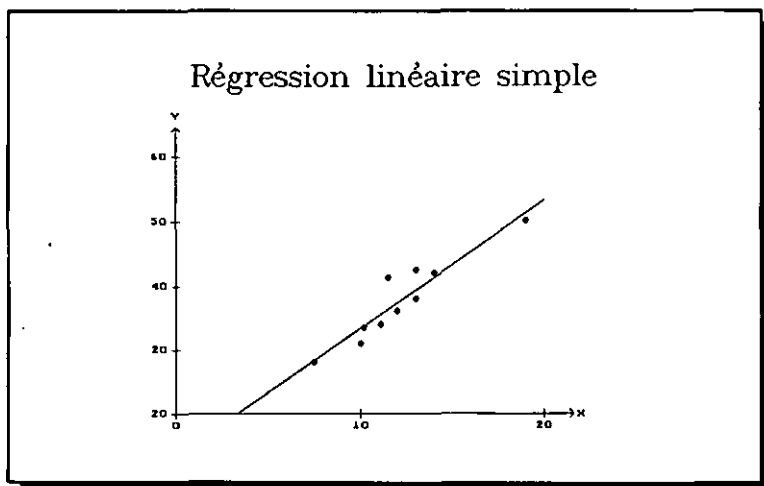


Figure 4.5: Outils statistiques: écran graphique complémentaire

4.3.3 Domaines couverts par les outils statistiques

Pratiquement, un outil statistique est un programme rattaché à une notion statistique particulière. Chaque outil statistique correspond donc à une des notions incluses dans le dictionnaire.

En fonction des types de résultats qu'ils permettent d'obtenir, nous pouvons classer ces outils en cinq catégories: les outils de calcul, les outils de représentation graphique, les outils d'analyse statistique, les outils d'accès aux tables statistiques et les outils de représentation des lois de probabilité.

1. Outils de calcul

Les outils de calcul correspondent à des notions dont le traitement n'amène qu'un résultat numérique simple, telles que la moyenne arithmétique, l'écart-type ou le coefficient de corrélation.

Ils offrent la possibilité d'entrer des données, d'obtenir le résultat numérique du traitement, de consulter la formule appliquée et d'obtenir des explications concernant les éventuels calculs intermédiaires.

Les données sont saisies dans le cadre situé sur la partie supérieure de la zone d'information de la page-écran standard. L'ensemble des résultats est affiché sur la partie inférieure de cette même zone.

2. Représentations graphiques

Les outils de représentation graphique correspondent à toutes les notions qui n'amènent pas de résultat numérique, mais une possibilité de représentation visuelle des données. L'histogramme, le box plot ou le diagramme en bâtons sont des exemples de représentations graphiques.

Ils permettent d'obtenir le graphe ou un diagramme correspondant aux données entrées.

L'entrée des données se fait sur la page-écran standard, et la représentation graphique occupe un écran graphique complet.

3. Outils statistiques d'analyse

Les outils d'analyse correspondent à l'ensemble des notions permettant d'étudier un ensemble de données de manière plus complexe. On y trouve par exemple l'analyse de variance, l'analyse de régression ou les différents tests d'hypothèses.

Ils peuvent offrir des résultats sous forme de tableaux ou de représentation graphique. Ils présentent également les conclusions à tirer des résultats de tests d'hypothèses.

4. Tables statistiques

Les outils d'accès aux tables statistiques permettent de consulter interactivement les tables susceptibles d'intervenir dans un des outils d'analyse statistique, telles que la table de la loi normale, de la loi de Student ou de la loi du chi-carré.

Ils permettent soit de consulter la table globale, soit d'effectuer des interrogations spécifiques sur le contenu de la table.

Lors de la consultation de la table globale, le système présente le contenu de la table sur un ou plusieurs écrans successifs, en fonction de la taille de la table. Lors d'une interrogation, le système n'affiche que le résultat relatif à la demande spécifique.

5. Lois de probabilité

Les outils de représentation des lois de probabilité discrètes ou continues offrent la possibilité de consulter l'allure des lois sur une image graphique, en fonction des paramètres de la loi choisis par l'utilisateur.

Les paramètres sont saisis sur l'écran standard, qui permet également de consulter la fonction de densité de la loi ainsi que les deux caractéristiques principales qui sont l'espérance mathématique et la variance. Un

second écran graphique fournit une représentation de la loi de probabilité en fonction des paramètres entrés.

4.4 Dictionnaire électronique

4.4.1 Objectifs du dictionnaire électronique

L'objectif du dictionnaire électronique est de mettre à disposition un "outil" permettant de consulter de manière interactive, rapide et efficace des informations relatives aux différents domaines de la statistique.

Le dictionnaire électronique répond à tout besoin ponctuel d'information. L'utilisateur doit donc être capable de déterminer quels sont ses besoins. Des index alphabétique et hiérarchique lui permettent de retrouver aisément l'information qu'il recherche.

Par le dictionnaire électronique, l'utilisateur a la possibilité d'accéder aux outils statistiques afin de traiter ses propres exemples, avec ses propres données. Un grand avantage pour l'utilisateur est donc de pouvoir participer activement à la compréhension des notions exposées dans le dictionnaire, les explications fournies par les outils statistiques étant intimement liées à l'exposé des méthodes dans le texte.

La présentation du dictionnaire sous forme de programme interactif présente de multiples avantages par rapport à la présentation sous forme d'un dictionnaire traditionnel. Nous pouvons mentionner notamment:

- la possibilité de "voyager" dans le dictionnaire de manière aisée, sans avoir à se préoccuper de rechercher la bonne page dans l'ouvrage: par une simple touche du clavier, l'utilisateur peut passer de la consultation d'un mot à un autre, accéder à un index alphabétique ou hiérarchique en vue de choisir le mot qui l'intéresse. L'utilisation de la souris permet également de sélectionner un mot directement dans le texte affiché à l'écran. L'ensemble de ces moyens permet à l'utilisateur de se déplacer sans effort de recherche, ce qui l'encourage à approfondir ses recherches d'information;
- la possibilité de revenir au mot précédemment consulté par une simple touche de fonction, ce qui évite à l'utilisateur de devoir tenir la page

d'un mot lorsqu'il veut en consulter un deuxième parallèlement;

- la possibilité de tester les notions expliquées dans le dictionnaire en entrant interactivement des données et en effectuant des calculs sur ces données.

4.4.2 Contenu du dictionnaire électronique

La base de données du dictionnaire électronique contient des informations concernant des notions statistiques, ou plus généralement des **mots**.

L'information concernant chaque mot est structurée en huit parties, ou huit **rubriques**:

1. Définition

Cette rubrique contient pour chaque notion statistique, une définition. Chaque notion statistique possède donc une entrée dans cette rubrique. Celle-ci constitue la rubrique principale de la base de données.

2. Historique

Cette rubrique possède une entrée pour chaque notion statistique pour laquelle il est possible d'établir un historique.

3. Aspects mathématiques

Cette rubrique contient pour chaque notion ayant une connotation mathématique une description détaillée des procédures de calcul.

4. Références

Dans cette rubrique, nous conservons pour chaque mot qui le nécessite, la liste des ouvrages cités en référence dans l'une ou l'autre des rubriques relatives au mot.

5. Domaines d'utilisation

Cette rubrique contient une explication des différentes possibilités d'utilisation de chaque notion.

6. Mots-clés

Cette rubrique offre une liste de mots-clés relatifs à chaque notion (voir section 3.8.3 pour une définition des mots-clés). Ceux-ci sont donnés en français, avec leur traduction anglaise.

7. Exemples

Cette rubrique possède une entrée pour chaque notion statistique susceptible d'être explicitée par un exemple.

8. Outil statistique

Cette dernière rubrique est une rubrique fictive puisqu'elle ne contient aucune information textuelle à consulter. C'est l'appel à cette rubrique qui offre l'accès aux outils statistiques exposés en section 4.3.

Il est à noter que la notion de **mot** doit être prise dans un sens très large. Cette notion peut représenter autant un mot de base tel que *échelle*, *classe*, *intervalle*, qu'une méthode d'analyse statistique telle que *test de Student* ou *analyse de variance*.

Nous gérons donc des données relatives:

- à des **mots**,
- et à des **rubriques**.

La figure 4.6 en donne une représentation schématique.

4.4.3 Dispositifs de dialogue

L'affichage par la machine des textes relatifs aux différents mots de la base de données se fait exclusivement par l'intermédiaire de la page-écran standard définie en section 4.2.

On peut constater sur la figure 4.7 que la zone de référence occupe les deux premières lignes de l'écran. La première ligne indique le module en cours ainsi que le mot consulté qui est ici la *moyenne arithmétique*, et la seconde ligne mentionne la rubrique en cours (*définition*) ainsi que la traduction anglaise du mot.

La zone d'information permet l'affichage d'une vue de 20 lignes du texte relatif à la rubrique consultée. Cette zone est définie comme un cadre déroulant, ce qui permet de faire défiler le texte de manière ascendante ou descendante.

La zone de commandes (dernière ligne de l'écran) indique en permanence la liste des touches de fonction actives. Celle-ci peut prendre alternativement deux

Dictionnaire

	Définition	Historique	Aspects mathématiques	Références	Domaines, limitations	Mots-clés	Exemples	Outils
Mots								

←———— Rubriques —————→

Figure 4.6: Représentation schématique du contenu du dictionnaire électronique

DICO Stat	MOYENNE ARITHMETIQUE
DEFINITION	ARITHMETIC MEAN
<p>La moyenne arithmétique est une mesure de tendance centrale particulière.</p>	
<p>Elle permet de caractériser le centre de la distribution de fréquences d'une variable quantitative en considérant toutes les observations et en leur attribuant le même poids (par opposition à la moyenne arithmétique pondérée).</p>	
<p>Elle est calculée par la somme des observations divisée par le nombre d'observations.</p>	
<p>F1:CHOIX MOT F2:MOT PRECEDENT F3:INDEX F4:CONTEXTE F5:IMPRIME --> ESC</p>	

Figure 4.7: Page-écran du dictionnaire électronique

formes différentes. La première, illustrée par la figure 4.7, permet d'effectuer des opérations relatives aux mots.

1. Choisir un mot

Le dictionnaire électronique ayant pour objectif principal de répondre à un besoin d'apprentissage ponctuel, il doit permettre d'accéder à n'importe quel mot de manière aisée.

Nous proposons différentes méthodes d'accès:

(a) Accès par un mot complet

Si l'utilisateur connaît la dénomination précise du mot auquel il désire accéder, il a la possibilité d'entrer le mot cherché et d'accéder ainsi directement à la rubrique *définition* correspondant au mot cherché.

(b) Accès par une image

Une image est un mot incomplet. La partie du mot qui est connue de l'utilisateur doit être mentionnée, et la partie inconnue peut être symbolisée par une étoile (*). Par exemple, 'test*' est l'image de tous les mots commençant par les lettres 'test'.

Lorsque l'utilisateur donne une image, il peut alors visualiser l'index alphabétique de tous les mots définis correspondant à cette image,

puis effectuer un choix parmi les mots proposés. La figure 4.8 en donne un exemple.

DICO Stat	
Liste des mots correspondant à votre choix	
1	TEST BILATERAL
2	TEST BINOMIAL
3	TEST D'ADEQUATION
4	TEST D'ADEQUATION DU CHI-CARRE
5	TEST D'HYPOTHESE
6	TEST D'INDEPENDANCE
7	TEST D'INDEPENDANCE DU CHI-CARRE
8	TEST DE FISHER
9	TEST DE KENDALL
10	TEST DE KOLMOGOROV-SMIRNOV
11	TEST DE KRUSKAL-WALLIS
12	TEST DE LA DIFFERENCE SIGNIFICATIVE MINIMALE
13	TEST DE MANN-WHITNEY
14	TEST DE SPERMAN
15	TEST DE STUDENT
16	TEST DE STUDENT POUR OBSERVATIONS PAIREES
17	TEST DE WILCOXON
18	TEST DE WILCOXON SIGNE

ESC

Figure 4.8: Index alphabétique

(c) Accès par un index alphabétique

Lorsque l'image se réduit simplement à une étoile, la liste des mots correspondant à cette image est alors la liste complète des mots du dictionnaire. Par ce moyen, l'utilisateur peut donc avoir accès à la liste alphabétique complète des mots définis.

(d) Accès par la souris

La souris, si elle est installée, offre un accès encore plus aisé. Elle permet de se déplacer sur l'écran et de sélectionner n'importe quel mot affiché à l'écran afin d'accéder à sa définition. Si le mot sélectionné n'existe pas dans le dictionnaire, le système affichera un message d'erreur. En vue d'éviter à l'utilisateur de sélectionner des mots qui aboutiront à un tel message, les mots accessibles sont affichés en couleur dans le texte.

2. Retourner au mot précédent

Cette possibilité de retour au mot précédemment consulté a pour but de faciliter la consultation parallèle de plusieurs mots. L'utilisateur peut à tout moment revenir au mot précédent par l'utilisation d'une simple touche de fonction, sans avoir à se rappeler de la dénomination exacte du mot puisqu'il ne doit pas le retaper.

3. Accéder à un index hiérarchique

Comme nous l'avons vu en section 3.8, les mots de la base de données peuvent être organisés selon une structure logique, en l'occurrence selon une structure hiérarchique. L'utilisateur a la possibilité de consulter cette structure par niveau et de choisir à chaque niveau la branche qu'il désire en vue de sélectionner le mot qui l'intéresse dans la structure. Comme l'illustre la figure 4.9, le premier niveau de la structure donne simplement les trois grands domaines statistiques inclus dans le dictionnaire.

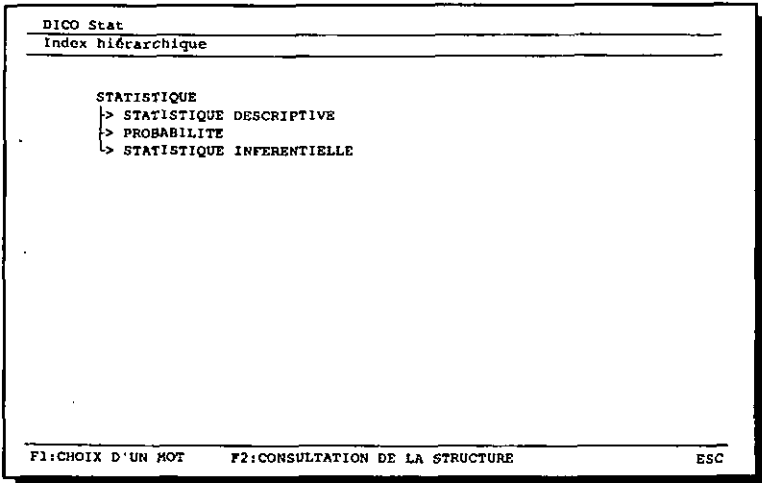


Figure 4.9: Index hiérarchique

4. Consulter le contexte dans lequel s'inscrit le mot

Ce contexte correspond à un positionnement du mot dans l'index hiérarchique. Un exemple est donné en figure 4.10 pour le contexte du mot *loi de probabilité continue*.

5. Imprimer le texte relatif au mot

Il est alors possible d'imprimer toutes les rubriques relatives au mot, ou uniquement une partie des rubriques laissée au choix de l'utilisateur.

La seconde forme de la zone de commandes offre la possibilité de passer de la consultation d'une rubrique à une autre, à l'intérieur d'un même mot. La figure 4.11 en propose un exemple.

Comme nous l'avons vu précédemment, toutes les rubriques ne contiennent pas forcément de l'information pour chaque mot. Seules les rubriques disponibles figureront donc sur la zone de commandes. La présence de la touche de fonction F8:OUTILS indique ici qu'un outil statistique est à disposition pour ce mot particulier.

4.5 Livre électronique

4.5.1 Objectifs du livre électronique

Le livre électronique est essentiellement un logiciel d'apprentissage. Son objectif est de regrouper l'essentiel de l'information contenue dans le dictionnaire électronique en un tout structuré, et de permettre ainsi au novice ou à toute personne ayant des lacunes dans un domaine particulier d'acquérir ou d'améliorer ses connaissances par la lecture d'un texte d'apprentissage structuré.

Le livre électronique est donc en quelque sorte un "guide" venant se greffer sur le dictionnaire. Ce guide amène une structure, un ordre dans l'apprentissage qui permet de répondre au besoin structuré d'information.

A tout moment lors de la lecture du texte, l'utilisateur peut accéder au dictionnaire, ce qui lui permet d'acquérir rapidement de l'information complémentaire sur une notion spécifique qu'il ne maîtrise pas.

Au cours de la lecture du livre, l'utilisateur peut également accéder aux outils statistiques:

- suite à une proposition du système qui suggère l'utilisation de chaque outil au moment opportun (par exemple, l'outil "analyse de variance" sera proposé lors de la consultation du chapitre y relatif);
- suite à une décision personnelle: l'utilisateur peut à tout moment, en cours d'apprentissage, accéder à n'importe quel outil statistique par l'intermédiaire du dictionnaire électronique. Il a ainsi la possibilité de vérifier pratiquement ses connaissances sur n'importe quelle méthode, même si celle-ci ne correspond pas au chapitre en cours de lecture.

La consultation du texte peut être faite de manière séquentielle (suite logique des chapitres du livre) ou dirigée. L'utilisateur a alors la possibilité de choisir

ses points d'entrée et d'organiser le flux de son apprentissage selon ses propres choix.

L'apport de l'outil informatique est un complément au livre traditionnel, qui permet à l'étudiant de devenir actif dans son apprentissage, de ressentir sa formation comme un jeu plutôt que comme un labeur. Il apporte notamment les avantages suivants:

- individualiser la formation, dans le sens où l'utilisateur (ou l'étudiant) a la possibilité d'investir son apprentissage dans les domaines qu'il désire;
- diminuer le temps de formation par rapport à un enseignement classique où la vitesse est souvent basée sur les étudiants les plus lents; l'étudiant a la possibilité d'exercer longuement les aspects qui posent des difficultés et de passer rapidement sur les points acquis.

4.5.2 Contenu du livre électronique

Le livre électronique est une méthode d'apprentissage des domaines traités dans le dictionnaire électronique. Le livre reprend en partie les textes du dictionnaire, ce qui peut paraître une redondance d'information. Toutefois, il faut bien être conscient de la divergence des buts poursuivis par les deux modules. Si la matière enseignée est la même, la façon dont l'information est structurée et transmise est différente.

La figure 4.12 donne une représentation schématique de la base de données du livre électronique. Elle est composée d'un certain nombre de chapitres, chaque chapitre étant subdivisé en sections et sous-sections.

Cette vision globale n'est toutefois pas suffisante. Il convient d'examiner de manière plus approfondie la façon dont est structurée chaque section. Comme le montre la figure 4.13, une section est composée de trois éléments:

- les pages **texte** qui constituent le corps de la section. Elles contiennent l'ensemble du texte relatif à la section (définitions, explications, exemples, etc);
- une ou plusieurs pages **résumé** qui contiennent un résumé du texte relatif à une section;

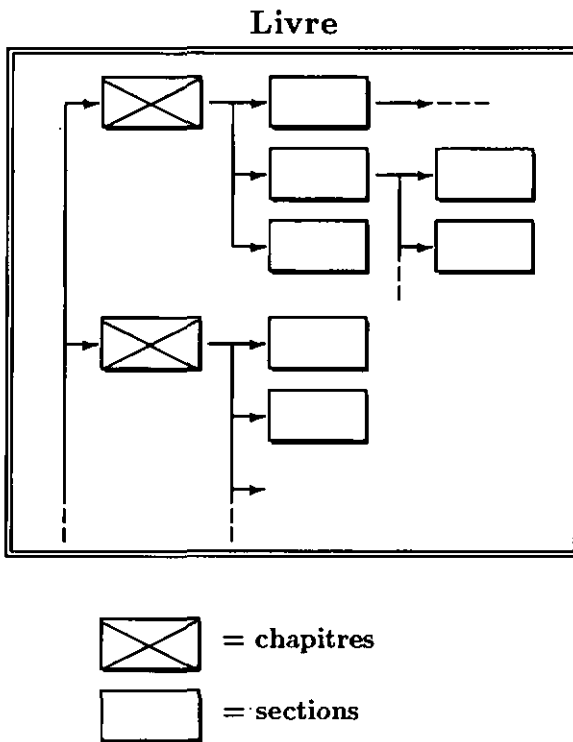


Figure 4.12: Représentation schématique du livre

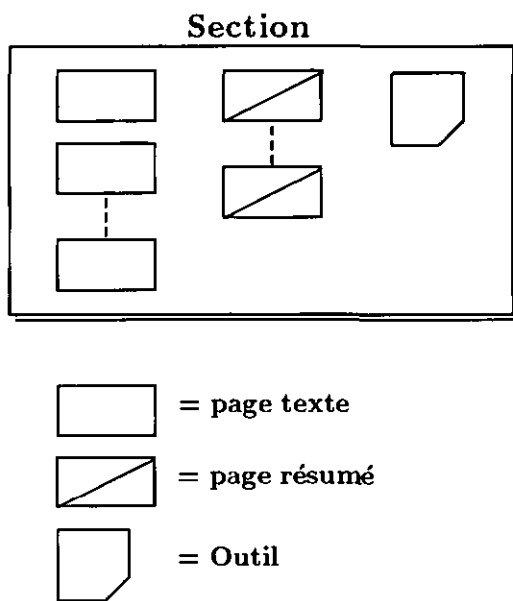


Figure 4.13: Représentation schématique d'une section

- éventuellement un **outil** qui est une possibilité d'accéder à un outil statistique particulier relatif au texte de la section.

4.5.3 Dispositifs de dialogue

En vue de répondre au critère d'homogénéité de la présentation de l'information entre les différents modules, c'est également la page-écran standard qui servira dans le livre électronique de dispositif central de communication entre la machine et l'utilisateur.

L'accès à l'apprentissage structuré peut se faire de plusieurs manières:

1. Accès par la table des matières

Il existe plusieurs façons de concevoir le parcours d'une table des matières:

- Une table des matières peut être parcourue de manière séquentielle. Toutefois, si la table est longue, il peut être rébarbatif de devoir parcourir toute la table pour choisir de consulter, par exemple, la dernière section.
- Une autre méthode est de concevoir la table des matières comme une structure hiérarchique qu'il est possible de consulter niveau par niveau. Dans ce cas, l'aperçu de la table est très fragmenté. On n'a pas de vision globale du contenu.

Nous avons choisi un compromis entre ces deux conceptions différentes, qui nous amène à consulter la table en deux temps:

- (a) dans un premier temps, comme l'illustre la figure 4.14, on affiche le premier niveau de la structure qui correspond à la liste des chapitres. L'utilisateur sélectionne le chapitre voulu simplement en se déplaçant dans cette liste, soit par l'utilisation des touches du clavier, soit plus simplement par la souris;
- (b) dans un second temps, lorsqu'un des chapitres a été sélectionné, on présente la "sous-table" complète relative à ce chapitre, ce qui permet un parcours séquentiel de l'ensemble des sections et sous-sections relatives au chapitre choisi. La figure 4.15 en donne un exemple.

LIVRE Stat	
Chapitre 1	Prologue..... 1
Chapitre 2	Population, variables et observations..... 17
Chapitre 3	Tableaux statistiques et représentations graphiques..... 49
Chapitre 4	Mesures de tendance centrale..... 151
Chapitre 5	Mesures de dispersion..... 235
Chapitre 6	Notions élémentaires de probabilités..... 310
Fin de la liste des chapitres - Sortie de la table des matières	
F1:Table	F2:Index F3:Continuation ESC

Figure 4.14: Table des matières du livre électronique (sommaire)

LIVRE Stat	
3	Tableaux statistiques et représentations graphiques..... 49
3.1	Introduction..... 49
3.2	Variables qualitatives..... 51
3.2.1	Répartition de population..... 52
3.2.2	Distribution de fréquences..... 54
3.2.3	Diagrammes en bâtons..... 58
3.2.4	Diagramme circulaire (Pie-chart)..... 67
3.3	Variables quantitatives discrètes..... 72
3.3.1	Distribution de fréquences cumulées..... 79
3.4	Variables quantitatives continues..... 83
3.4.1	Organisation par classes..... 84
3.4.2	Histogrammes..... 88
3.4.3	Les polygones et les courbes de fréquences..... 114
3.5	Considérations similaires pour deux variables..... 123
3.5.1	Tableaux croisés..... 125
3.5.2	Représentation graphique..... 142
Fin de la liste des sections - Retour à la liste des chapitres	
F1:Table	F2:Index F3:Continuation ESC

Figure 4.15: Table des matières du livre électronique (détail)

Avec un tel compromis, on conserve une vision globale du contenu puisque la liste des chapitres est présentée en bloc. L'aperçu de la table est fragmenté, mais dans une limite raisonnable puisque l'ensemble des sections relatives à un chapitre peut également être consulté en bloc. Ces blocs n'étant pas trop grands, l'aspect rébarbatif du parcours séquentiel d'une longue liste est également évité.

2. Accès par l'index

L'index est une liste alphabétique de mots. On associe à chaque mot une liste de numéros de pages auxquelles il peut être utile de se référer. L'index permet donc d'accéder directement aux endroits ponctuels dans le livre où la notion à laquelle on s'intéresse est utilisée.

L'accès par l'index se fait également en deux temps:

- (a) dans un premier temps, comme on peut le voir sur la figure 4.16, on affiche la liste alphabétique des mots de l'index;

LIVRE Stat			
Index			
A			
algorithme			
algorithme de Yates			
analyse combinatoire			
analyse de données			
analyse de variance			
analyse des résidus			
analyse factorielle			
arrangement			
B			
base			
biais			
binôme			
box plot			
C			
catégorie			

F1: Recherche lettre	F2: Recherche chaîne	F3: Chaîne suivante	ESC

Figure 4.16: Index du livre électronique (sommaire)

- (b) dans un second temps, après qu'un des mots ait été sélectionné, on présente la liste des numéros de page associés à ce mot. La figure 4.17 en donne une illustration.

3. Continuation de la lecture

Cette option n'est valable que lors de l'entrée dans le livre, contrairement

LIVRE Stat	
Index	
A	
algorithme	
algorithme de Yates	
analyse combinatoire	
analyse de données	
analyse de variance	
analyse des résidus	
analyse factorielle	
arrangement	
B	
base	
biais	
binôme	
box plot	
C	
catégorie	
Liste des pages	
6	(6/16) Prologue
312	(3/57) Notions élémentaires de probabilité
317	(8/57) Notions élémentaires de probabilité
356	(47/57) Notions élémentaires de probabilité
F1: Recherche lettre F2: Recherche chaîne F3: Chaîne suivante ESC	

Figure 4.17: Index du livre électronique (détail)

aux deux précédentes qui sont accessibles à n'importe quel moment lors de la lecture du texte.

Elle permet de retrouver automatiquement la dernière page consultée lors de la dernière utilisation du livre électronique. Lors de la première consultation du livre, cette option donne accès à la première page du livre.

Ces différentes possibilités d'accès offrent une structure d'apprentissage qui permet à l'utilisateur qui connaît exactement dans quels domaines il désire acquérir des connaissances de commencer son apprentissage à partir de n'importe quel domaine, et de parcourir les différents domaines dans l'ordre qu'il juge favorable.

Pour l'utilisateur qui n'a aucune notion, ce problème de choix peut être évité en suivant l'ensemble du texte dans l'ordre logique qui est proposé.

La figure 4.18 présente la consultation d'une page de texte standard. La zone de référence indique sur la première ligne le numéro et le titre du chapitre en cours, ainsi que le numéro de la page consultée. Les numéros entre parenthèses indiquent la position de la page dans le chapitre. Par exemple, (69/75) indique que l'on se trouve à la 69ème page du chapitre, qui en compte 75. Sur la seconde ligne figurent le numéro et le titre de la section, ainsi que le type de

page consultée (texte ou résumé).

LIVRE Stat	5 Mesures de dispersion	Page 303 (69/75)
5.5.1 Variables dichotomiques		(Texte)

Les valeurs prises par une variable dichotomique peuvent être associées aux nombres 0 et 1. Par exemple, pour la variable dichotomique "sexe" dont les catégories sont H pour homme et F pour femme, on peut associer le chiffre 0 à la catégorie H et le chiffre 1 à la catégorie F. Dans ce sens, la variable qualitative "sexe" peut donc être représentée par une variable quantitative X prenant les valeurs numériques 0 et 1,

$$X = \begin{cases} 0 & \text{si sexe} = H \\ 1 & \text{si sexe} = F \end{cases}$$

Les observations x_1, \dots, x_n représentent donc le sexe, homme ou femme d'un ensemble de n personnes.

F1:Table	F2:Index	F3:Dico	F4:Imprime	PgDn	PgUp	ESC
----------	----------	---------	------------	------	------	-----

Figure 4.18: Page-écran du livre électronique

La zone de commande indique les différentes possibilités de l'utilisateur:

- **Accès à la table des matières**

La consultation, en cours de lecture, de la table des matières offre à l'utilisateur la possibilité de continuer sa lecture à partir de l'endroit de son choix. Il peut ainsi par exemple abrégier la lecture consécutive d'un chapitre ou d'une section afin de passer directement au chapitre suivant ou à la section suivante.

- **Accès à l'index**

L'accès à l'index décrit précédemment est la seconde possibilité mise à disposition de l'utilisateur en vue de modifier le parcours séquentiel de sa lecture.

- **Accès au dictionnaire électronique**

L'accès au dictionnaire peut se faire à n'importe quel moment, en vue de rechercher n'importe quel mot. Ceci permet d'obtenir rapidement des informations plus complètes sur toute notion particulière du livre, ce qui peut être très utile afin d'avoir une vision regroupée de toute l'information relative à un mot spécifique, ou également pour accéder à un outil statistique particulier.

- **Impression**

Cette fonction d'impression permet d'obtenir une copie sur papier de la totalité du chapitre en cours.

- **Accès à un outil statistique**

Lorsque le contexte s'y prête, cette fonction permet d'accéder directement à un outil statistique spécifique correspondant à la matière en cours. Il est alors possible de tester interactivement la compréhension de la matière traitée à un moment donné.

- **Passer à la page suivante ou revenir à la page précédente**

Ce sont les deux touches de fonction PgUp (page précédente) et PgDn (page suivante) qui permettent de parcourir le livre séquentiellement page par page, et de revenir en cas de besoin aux pages lues antérieurement.

Partie II

ANALYSE INFORMATIQUE

Chapitre 5

Analyse informatique: architecture globale

5.1 Introduction

Cette deuxième partie est consacrée à la présentation de notre application du point de vue du développement informatique.

Comme nous l'avons déjà vu en section 2.7, l'application globale considérée comprend trois éléments fonctionnels principaux dont l'articulation a été décrite et représentée en figure 2.2:

- le dictionnaire électronique
- le livre électronique

- les outils statistiques

Bien qu'il existe certains liens permettant un passage entre ces trois éléments, ils peuvent être considérés et analysés en tant que trois modules informatiques indépendants.

Nous avons, dans le chapitre précédent, donné une description de l'application globale, ce qui nous a amené à présenter, pour chacun de ces éléments, les différents dispositifs de dialogue relatifs à l'interface utilisateur.

L'interface utilisateur n'est toutefois qu'une partie de l'environnement complet dans lequel s'inscrit toute application informatique. L'objectif de ce chapitre est de présenter brièvement l'environnement relatif à chaque module: les deux premiers sont des modules qui s'inscrivent dans l'environnement d'une application de consultation. Le troisième s'inscrit dans l'environnement d'une application interactive de calcul.

5.2 Application de consultation

Une application de consultation est un outil par lequel la machine fournit des informations à l'homme, le déclenchement de l'affichage d'une information particulière pouvant provenir soit d'un ordre donné par l'homme, soit d'une règle connue de la machine. L'ensemble des informations consultables est stocké dans une **base de données** qui constitue le noyau central d'une telle application.

La création d'une application de consultation implique d'une part la possibilité de remplir la base de données, et d'autre part la possibilité de la consulter.

L'environnement dans lequel s'inscrit une telle application est représenté en figure 5.1.

Il est caractérisé par une double interface:

- une **interface opérateur**:

l'opérateur est la personne responsable de la création et de la maintenance de la base de données. Sa tâche première est la saisie des textes de base destinés à être inclus dans la base de données. Il en gère ensuite le contenu par l'intermédiaire d'une interface qui lui est proposée par le programme de génération de la base de données;

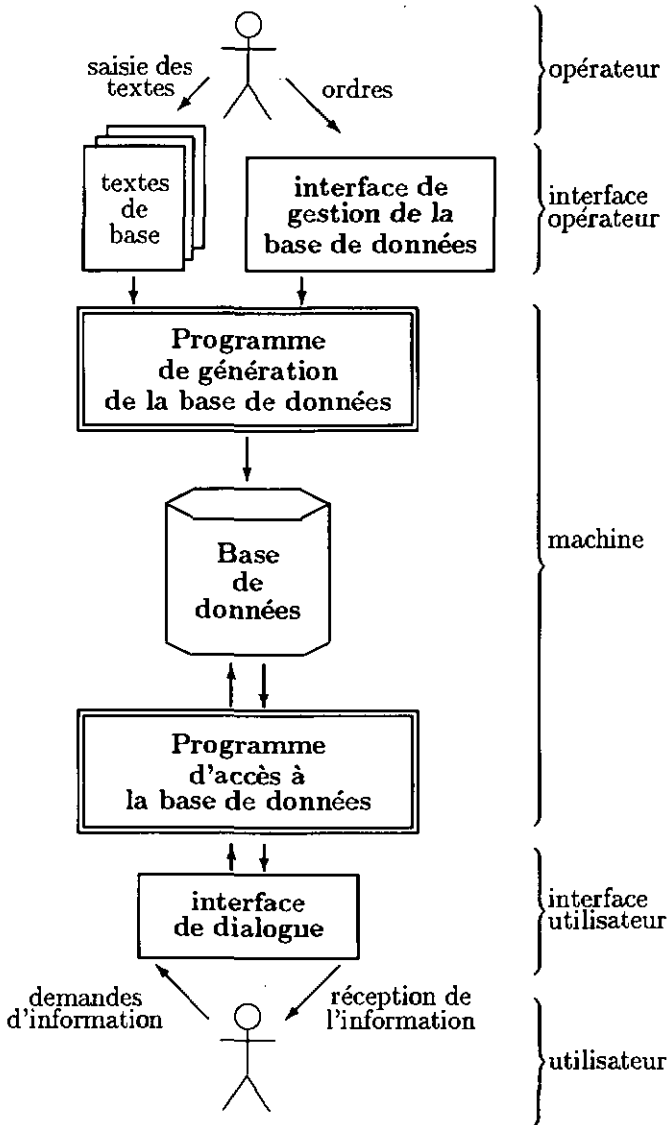


Figure 5.1: Environnement d'une application de consultation

- une **interface utilisateur**:

l'utilisateur n'a qu'un droit de consultation sur la base de données. Il a donc la possibilité de demander et de recevoir de l'information, au travers de l'interface de dialogue qui lui est fournie par le programme d'accès à la base de données.

Que nous parlions de dictionnaire ou de livre électronique, le problème est dans les deux cas la consultation d'une base de données: dans le premier cas, celle-ci est composée de mots pour lesquels l'information est ventilée en différentes rubriques, alors que dans le second elle est composée d'un ensemble de chapitres subdivisés en sections et sous-sections.

La structure de la base de données est donc différente dans les deux cas, de même que les modalités d'accès.

Nous pouvons distinguer le dictionnaire du livre électronique en fonction des principaux critères suivants:

	Dictionnaire électronique	Livre électronique
Base de données	structure à plat (matricielle)	structure hiérarchique
Principale méthode d'accès	direct par un mot	direct par un chapitre, section ou sous-section
Nature du texte	chaque mot est considéré comme un tout homogène et cohérent	l'ensemble du livre doit former un tout homogène et cohérent
Consultation possible	consultation ponctuelle d'un mot (une consultation séquentielle n'a pas de sens)	consultation séquentielle de l'ensemble du livre, ou d'un sous-ensemble particulier

Dans la mesure où ni le flux, ni la structure, ni la nature de l'information véhiculée ne sont les mêmes dans le dictionnaire et dans le livre, il semble judicieux de prévoir deux bases de données de structure différente et possédant chacune des méthodes d'accès personnalisées.

Bien que la structure de l'environnement reste le même pour les deux appli-

cations, chaque élément de cette structure est donc à redéfinir dans chaque cas.

5.3 Application interactive de calcul

Une application interactive de calcul est un outil par lequel l'homme fournit des informations (entrées) à la machine, informations qui seront utilisées et traitées par la machine afin de présenter les résultats du traitement (sorties) à l'homme. Ces résultats peuvent être de différentes formes:

- reformulation des entrées (présentation de données brutes sous la forme d'un tableau de fréquences),
- résumé des entrées (affichage de statistiques telles que la moyenne ou l'écart-type),
- résultats d'une analyse appliquée aux données (équation de régression, résultats d'un test d'hypothèse),
- représentation graphique (dessin d'un graphique de dispersion, d'une distribution de fréquences).

En fonction des résultats fournis, l'homme peut alors modifier ses entrées et provoquer de nouveaux traitements.

L'environnement d'une telle application est schématisé en figure 5.2. Le noyau central n'est plus ici une base de données, mais un ensemble de routines regroupées en un **programme de calcul**. Cet environnement n'est caractérisé que par une seule interface: l'interface utilisateur. Celle-ci joue ici un double rôle puisque l'utilisateur est à la fois le "donneur" et le "receveur" d'information.

5.4 Outils de développement

Nous présentons dans les chapitres suivants une analyse plus approfondie de chacun des trois modules (dictionnaire électronique, livre électronique et outils statistiques).

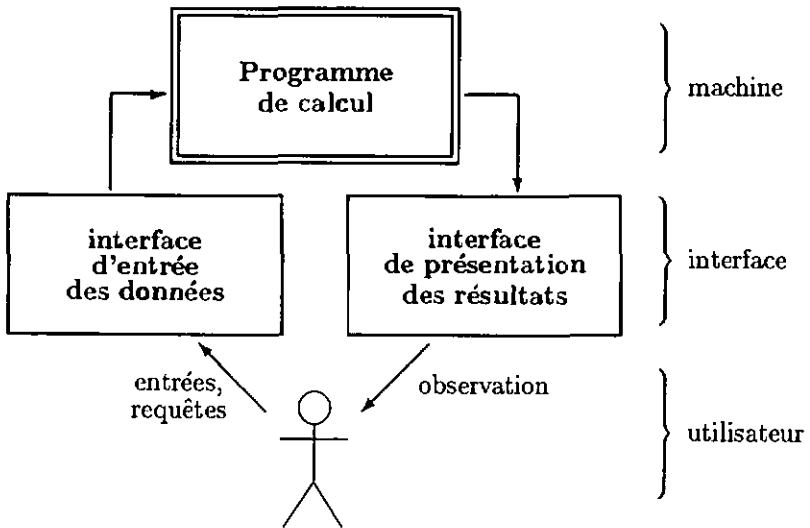


Figure 5.2: Environnement d'une application interactive de calcul

Celle-ci est liée tout au long du texte aux outils (matériel et logiciel) de développement que nous avons utilisés pour la construction de notre application.

Nous avons choisi de développer l'application globale sur un *Personal System* (IBM PS/2), assurant ainsi la portabilité de l'application sur tous les PC (*Personal Computer*) IBM et compatibles.

Au niveau du langage de programmation, nous avons choisi d'utiliser Turbo Pascal (version 5.0) [60,61] pour le développement de la totalité des modules, celui-ci offrant d'une part un environnement de développement très agréable (éditeur intégré, détection interactive des erreurs, débogueur intégré), et d'autre part un grand nombre de fonctionnalités intéressantes:

- un langage structuré;
- des routines standard de gestion de l'affichage à l'écran (gestion des fenêtres, des couleurs), du haut-parleur interne et du clavier;
- un mécanisme de gestion des partiels (parties de programme partageant

tour à tour une même zone de mémoire) offrant la possibilité de faire exécuter un programme dont la taille est supérieure à la mémoire disponible;

- un module de gestion de l'écran en mode graphique (texte et dessins);
- un accès facilité à de nombreuses fonctions du DOS (recherches dans le système de répertoires, lancement d'autres programmes).

Chapitre 6

Analyse informatique du dictionnaire électronique

6.1 Introduction

Le dictionnaire électronique est, selon la terminologie présentée au chapitre précédent, une application de consultation. Nous présentons dans ce chapitre les différents éléments de cet environnement de développement.

Nous commencerons par la description de la structure de la base de données. Nous décrivons ensuite le programme de génération de cette base de données ainsi que l'interface opérateur s'y rapportant. Nous proposons ensuite de développer les différentes méthodes d'accès à la base de données proposées par le programme d'interrogation du dictionnaire électronique.

6.2 Structure de la base de données

Le dictionnaire électronique permet à l'utilisateur de consulter des informations spécifiques (rubriques) relatives à des notions (mots).

Comme le montre la figure 6.1, nous gérons une information qui peut être structurée sous une forme matricielle à deux dimensions. La première dimension est constituée de l'ensemble des mots du dictionnaire, et la seconde des rubriques de texte relatives à chaque mot.

		RUBRIQUES						
		1	2	3	4	5	6	7
M O T S	M							
	O							
	T							
	S							

Figure 6.1: Représentation matricielle de la base de données du dictionnaire électronique

Appelons cellule une case de cette matrice. Chaque cellule est caractérisée par deux appartenances: l'appartenance à un mot et l'appartenance à une rubrique. Une cellule est composée d'un certain nombre de lignes de texte.

Examinons différentes possibilités de stockage de l'information contenue dans cette matrice.

1. Stockage des cellules de taille fixe

La possibilité de stockage la plus simple est de fixer la taille d'une cellule à un nombre fixe de lignes, et de stocker une matrice de dimension

$$N \times L \times 7,$$

où N est le nombre total de mots définis dans le dictionnaire, L le nombre de lignes réservées par cellule, et 7 le nombre de rubriques.

Nous pouvons remarquer que dans l'élaboration des textes, le nombre de lignes dans chaque cellule est variable. De plus, certaines cellules peuvent être vides puisque seules les cellules correspondant à la première rubrique sont obligatoirement non vides.

Notre expérience nous a montré que le nombre de lignes moyen par cellule dépend très fortement de la rubrique. Sur la base des textes que nous avons inclus dans notre application, nous avons déterminé ce nombre moyen de lignes par cellule (non vide) pour chaque rubrique. Le tableau 6.1 donne également le pourcentage de mots ayant une entrée dans chaque rubrique (pourcentage de cellules non vides):

	Rubrique	Nombre moyen de lignes par cellule	Pourcentage de cellules non vides
1	Définition	12.1	100%
2	Historique	11.7	54%
3	Aspects mathématiques	47.0	65%
4	Références	9.5	50%
5	Domaines et limitations	18.8	52%
6	Mots-clés	7.8	95%
7	Exemples	56.6	60%

Table 6.1: Taille des cellules du dictionnaire

Au niveau de la taille moyenne des cellules, nous pouvons constater que les rubriques 4 (références) et 6 (mots-clés) sont les plus courtes. Les rubriques 1 (définition), 2 (historique) et 5 (domaines et limitations) sont également relativement courtes. Les deux rubriques qui occupent le plus grand nombre de lignes par cellule sont la rubrique 3 (aspects mathématiques), et surtout la rubrique 7 (exemples).

Au niveau du pourcentage de cellules non vides, nous pouvons remarquer qu'excepté la rubrique 1 (définition) qui est présente pour chaque mot, le taux de cellules non vides par rubrique varie entre 50% (références) et 95% (mots-clés).

Ces quelques remarques montrent bien que le stockage de cellules de taille fixe est loin d'être optimal, car l'espace disque nécessaire au stockage de la matrice est bien supérieur à l'espace véritablement utilisé. En tenant compte que la cellule la plus longue comprend 300 lignes, et qu'une ligne

est composée de 80 caractères, nous pouvons estimer cet espace. Le tableau 6.2 présente les calculs de l'espace disque nécessaire, ainsi que de l'espace utilisé, basé sur les données du tableau 6.1.

espace disque nécessaire par mot	espace disque moyen utilisé par mot
300 lignes ×80 caractères ×7 rubriques	$[(12.1 \times 1) + (11.7 \times 0.54) + (47.0 \times 0.65) + (9.5 \times 0.5) + (18.8 \times 0.52) + (7.8 \times 0.95) + (56.6 \times 0.6)] \times 80$ caractères
= 168'000 caractères	= 8'389.12 caractères

Table 6.2: Espace disque nécessaire au stockage de cellules de taille fixe

Nous constatons qu'avec un stockage de cellules de taille fixe, 4.99% uniquement de l'espace occupé serait utilisé au stockage de l'information!

2. Stockage de cellules de taille variable

La méthode de stockage que nous avons choisie est indépendante de la taille des cellules.

Considérons l'ensemble des cellules relatives à la première rubrique. Les lignes de texte relatives à chacune de ces cellules seront stockées séquentiellement dans un fichier appelé DICO1. La figure 6.2 présente un exemple du fichier DICO1. Celui-ci est constitué de 3 mots comprenant respectivement 3, 4 et 5 lignes de texte.

Les rubriques 2 à 7 sont stockées de la même manière dans les fichiers DICO2 à DICO7. Les fichiers de base sont donc au nombre de 7:

- (a) DICO1: fichier de la rubrique *définition*
- (b) DICO2: fichier de la rubrique *historique*
- (c) DICO3: fichier de la rubrique *aspects mathématiques*
- (d) DICO4: fichier de la rubrique *références*
- (e) DICO5: fichier de la rubrique *domaines et limitations*
- (f) DICO6: fichier de la rubrique *mots-clés*
- (g) DICO7: fichier de la rubrique *exemples*

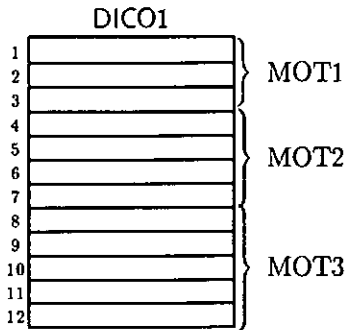


Figure 6.2: Fichier de base du dictionnaire électronique

De cette manière, l'espace disque utilisé correspond exactement à l'espace nécessaire au stockage de l'information, indépendamment de la variabilité de la taille des cellules. De plus, ceci offre également l'avantage de ne pas devoir fixer arbitrairement une taille de cellule maximale.

Il reste à conserver les adresses permettant de retrouver dans ces fichiers les lignes de texte relatives à un mot. Ces adresses sont conservées dans un fichier INDEX. Celui-ci est composé d'enregistrements que nous appellerons `Index_Enre`. Chaque enregistrement comporte l'information relative à un mot et est composé de trois champs. La description en Turbo Pascal d'un enregistrement est la suivante:

```

Index_Enre : Record
    Mot           : String[60]
    Pointeur_Debut : Array[1..8] Of Integer
    Pointeur_Fin  : Array[1..8] Of Integer
End

```

Les champs `Index_Enre.Pointeur_Debut` et `Index_Enre.Pointeur_Fin` sont composés de huit positions. Les sept premières positions donnent respectivement le numéro de la première et de la dernière ligne de texte relative au mot dans le fichier `DICOx`, où `x` est le numéro de la position. Les valeurs possibles pour ces champs sont:

- le numéro de ligne correspondant au début, respectivement à la fin de la cellule dans le fichier `DICOx` si l'entrée dans la cellule existe;
- la valeur `-1` si le mot n'a pas d'entrée dans la cellule.

La huitième position indique uniquement si l'accès aux outils statistiques est possible ou non pour chaque mot.

Reprenons notre exemple de la figure 6.2 afin de comprendre de quelle manière les informations nécessaires à l'accès aux fichiers DICOx sont stockées dans le fichier INDEX.

Le premier mot MOT1 occupe les trois premières lignes du fichier DICO1. Ces informations se retrouvent dans l'enregistrement relatif à MOT1 de l'INDEX: la première position du champ `Index.Enre.Pointeur.Debut` indique le numéro de ligne correspondant au début du mot dans la première rubrique (1ère ligne), et la première position du champs `Index.Enre.Pointeur.Fin` indique le numéro de ligne correspondant à la fin du mot dans la même rubrique (3ème ligne).

Pour le deuxième mot MOT2, les valeurs de la première position des champs `Index.Enre.Pointeur.Debut` et `Index.Enre.Pointeur.Fin` seront donc respectivement égales à 4 et 7 puisque MOT2 occupe les lignes 4 à 7 du fichier DICO1.

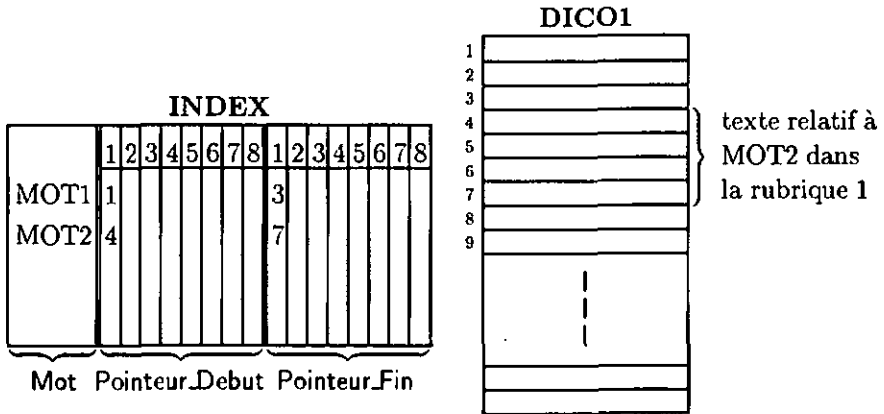


Figure 6.3: Accès aux fichiers DICOx par l'index

La figure 6.3 présente schématiquement les informations relatives à cet exemple. Si l'on connaissait le nombre de lignes relatives à chacun de ces mots dans les autres rubriques, on pourrait ainsi déterminer la valeur des positions suivantes des champs `Index.Enre.Pointeur.Debut` et `Index.Enre.Pointeur.Fin` pour chaque mot.

Afin de permettre un accès rapide aux fichiers DICOx, le fichier INDEX est automatiquement trié par ordre alphabétique des mots.

6.3 Description du programme de génération de la base de données

La structure de la base de données étant définie, il est nécessaire à présent de déterminer quels moyens nous allons mettre à disposition afin de pouvoir y introduire le contenu (le texte relatif à chaque rubrique de chaque mot).

Nous sommes en présence d'un processus dans lequel nous trouvons en amont un ensemble de textes sous une structure quelconque, et en aval les mêmes textes structurés selon les critères spécifiques à la base de données.

Ce processus est réalisé concrètement par un programme que nous appellerons **GENDICO** (programme de GENÉration de la base de données du DICtiOnnaire). Celui-ci a pour but de gérer la structure et le contenu de la base de données destinée à être consultée par le programme d'interrogation du dictionnaire.

Comme le montre la figure 6.4, le programme **GENDICO** accepte en entrée des fichiers de texte. Il fournit en sortie les fichiers de la base de données qui ont été décrits dans la section précédente: DICO1 à DICO7, ainsi que le fichier INDEX.

On peut voir sur la même figure que **GENDICO** permet également de créer deux fichiers supplémentaires: les fichiers ACCES et ARBRE qui seront décrits ci-dessous.

L'ensemble des tâches du programme **GENDICO** peut être divisé en trois opérations principales que nous allons étudier de manière plus détaillée:

- la génération des fichiers de base;
- la gestion des accès;
- la gestion de l'index hiérarchique.

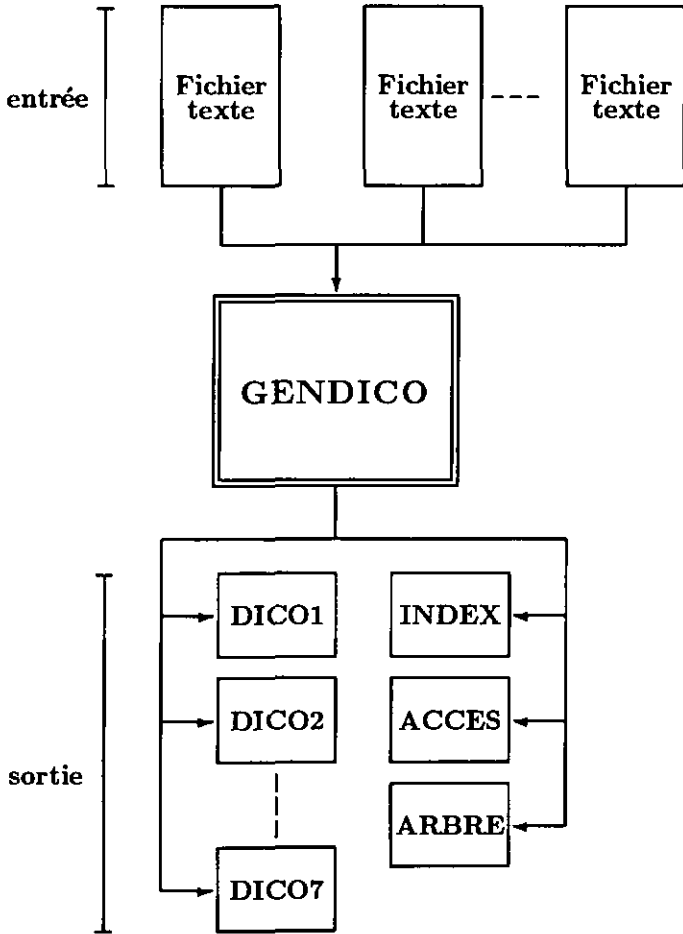


Figure 6.4: Processus de génération du dictionnaire électronique

6.3.1 Génération des fichiers de base

Le processus de génération des fichiers de base consiste à créer, à partir de fichiers textes entrés en amont du processus,

- les fichiers DICO1 à DICO7;
- et le fichier INDEX y permettant l'accès.

Les fichiers texte doivent avoir été saisis selon des critères de formatage spécifiques. Ces fichiers doivent être structurés de telle manière que le programme de génération de la base de données puisse notamment repérer:

- le début et la fin de chaque cellule;
- le mot traité;
- la rubrique traitée.

La figure 6.5 montre la structure globale d'un tel fichier.

Les règles de saisie du texte relatif à une cellule sont les suivantes:

1. règles relatives à une cellule:

le texte relatif à une cellule doit obligatoirement figurer dans un seul fichier. Une cellule relative au mot MOT doit être précédée d'une ligne standard:

..x MOT ,

où x est le numéro de la rubrique. La fin de la cellule doit être indiquée par une seconde ligne standard:

..PAGE .

2. règles relatives aux graphiques:

il est possible à tout endroit dans le texte de faire appel à une image graphique. Lorsque des graphiques doivent être introduits dans le texte, ceux-ci doivent avoir été préalablement créés et sauvegardés dans un fichier (par exemple GRAPHE.PIC).

Fichier texte

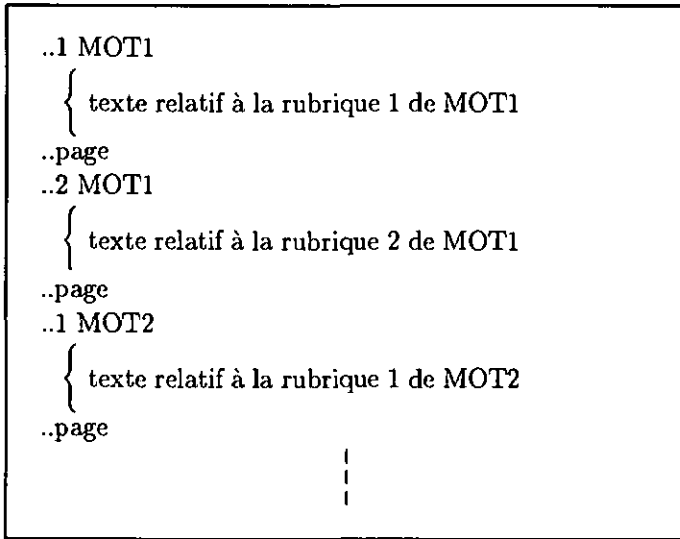


Figure 6.5: Structure d'un fichier texte (input) du dictionnaire

Pour faire appel à ce graphique, il est nécessaire d'entrer dans le texte une information indiquant à l'utilisateur qu'il pourra consulter une image en tapant une touche de fonction. La touche <F9> a été retenue pour ce type d'opération. Par exemple:

Tapez <F9> pour construire
le graphique

Dans les textes de base, il sera nécessaire d'indiquer sur la ligne précédant ce message le nom du fichier contenant l'image graphique. Ce nom doit être précédé d'un caractère spécial afin que le programme de consultation puisse le reconnaître et ne pas l'afficher à l'écran. Nous avons retenu le caractère ASCII 31 pour ce type d'indication, ce caractère n'étant habituellement pas utilisé dans la rédaction des textes.

3. règles relatives aux mots du dictionnaire:

lors de l'interrogation du dictionnaire électronique, l'utilisateur aura la possibilité de passer de la consultation d'un mot à un autre en sélectionnant une touche de fonction.

tionnant avec la souris le mot qui l'intéresse dans le texte. Nous avons mentionné au chapitre 4 que les mots qui possèdent une entrée dans le dictionnaire apparaissent en couleur dans le texte affiché à l'écran.

Ces mots seront reconnaissables par le programme d'interrogation du dictionnaire qui pourra alors les mettre en évidence lors de l'affichage du texte s'ils sont entourés du caractère ASCII 18, caractère qui n'est également pas utilisé dans la rédaction habituelle de textes.

4. règles relatives aux sauts de pages:

lors de l'impression du contenu du dictionnaire, il est nécessaire que le système soit capable de détecter les endroits où il lui est permis de passer à une page suivante. Il ne doit pas par exemple passer à une page suivante au milieu d'une formule ou d'un tableau, ou juste après un titre.

Pour ce faire, il est nécessaire de faire précéder toute portion de texte qui ne doit pas être coupée par le caractère ASCII 21, et de la faire suivre par le caractère ASCII 22. Ces caractères doivent obligatoirement se trouver en première colonne dans le fichier (premier caractère de la ligne).

6.3.2 Gestion des accès

La gestion des accès permet de créer et de gérer le fichier ACCES. Celui-ci n'est pas créé automatiquement par GENDICO mais doit être géré par l'opérateur.

Il contient des informations relatives aux mots qui ne sont pas saisies directement dans les fichiers de base:

1. il stocke d'une part les équivalents anglais de chaque mot français figurant dans l'INDEX; la saisie des mots anglais est facultative, mais elle est très utile dans la consultation du dictionnaire électronique, car elle permet de voir s'afficher la traduction anglaise dans la zone de référence lors de la consultation d'un mot, et d'offrir un accès multilingue aux informations contenues dans le dictionnaire;
2. il conserve d'autre part un numéro d'accès aux outils statistiques.

Chaque enregistrement `Acces.Enre` de ce fichier contient trois champs dont la description en Turbo Pascal est la suivante:

```
Acces.Enre : Record
                Mot_Anglais      : String[60]
                Mot_Français     : String[60]
                Numero_Logiciel  : Integer
            End
```

Le fichier `ACCES` contient autant d'enregistrements que l'`INDEX`. Pour chaque mot `Acces.Enre.Mot_Français`, son équivalent anglais est stocké dans `Acces.Enre.Mot_Anglais` et le numéro d'accès aux outils statistiques dans `Acces.Enre.Numero_Logiciel`. (Le chapitre 8 donne plus de détails concernant la gestion des outils statistiques.)

De même que le fichier `INDEX`, `ACCES` est automatiquement trié dans l'ordre alphabétique de son premier champ.

6.3.3 Gestion de l'index hiérarchique

Comme nous l'avons défini en section 4.4.3, le dictionnaire électronique permet d'accéder aux différents mots par l'intermédiaire d'un index hiérarchique.

GENDICO offre les moyens nécessaires à la création et à la gestion de cet index au travers d'un programme de gestion d'"arbres". Celui-ci permet de créer et de mettre à jour le fichier `ARBRE` contenant l'index hiérarchique.

La création de cet index n'est pas obligatoire. Toutefois, son existence fournit, lors de la consultation du dictionnaire, des possibilités intéressantes en permettant d'une part d'accéder aux mots de manière plus parlante que par un simple index alphabétique, et d'autre part de pouvoir visualiser le contexte relatif à chaque mot.

Un arbre est composé de plusieurs éléments:

- une **racine**: le sommet de l'arbre
- des **feuilles**: les éléments terminaux
- des **noeuds**: les éléments intermédiaires

- des **branches**: les relations

Tout élément de l'arbre possédant une structure inférieure est appelé un **père**. Ses éléments inférieurs sont appelés ses **fil**s. L'ensemble des fils du même père sont dits **frères**. Ces différents éléments sont représentés en figure 6.6.

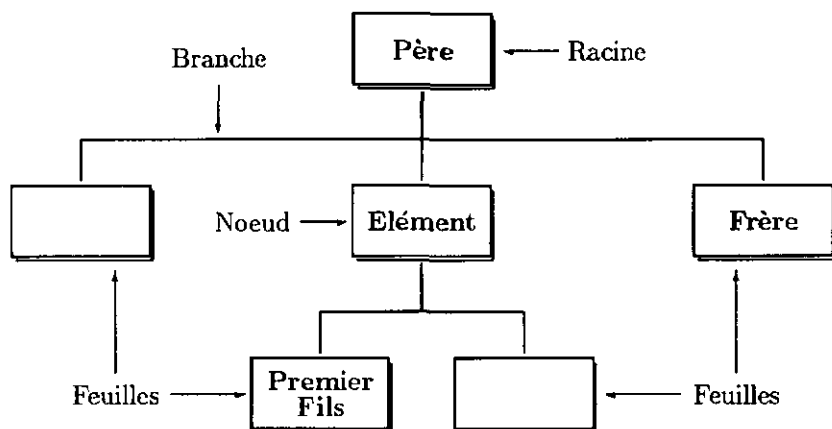


Figure 6.6: Terminologie associée à une structure hiérarchique

Le fichier ARBRE contient les enregistrements `Arbre.Enre` permettant de représenter l'ensemble de la structure hiérarchique de notre application.

Chaque enregistrement correspond à un mot. A chaque mot est associé un numéro unique, de même que son premier fils, son frère et son père. La description en Turbo Pascal d'un enregistrement est la suivante:

```
Arbre.Enre = Record
    No_Element : Integer
    Mot        : String[60]
    Premier_Fils : Integer
    Frere      : Integer
    Pere       : Integer
End
```

6.3.4 Interface opérateur

Le premier travail de l'opérateur est la saisie des textes en fonction des critères mentionnés en section 6.3.

Une fois les textes saisis, **GENDICO** offre à l'opérateur une interface de travail lui permettant d'effectuer les manipulations nécessaires à la création et à la gestion de la base de données. Comme le montre la figure 6.7, l'interface se présente sous forme de menus qui offrent à l'opérateur les possibilités suivantes:

1. **Destruction de la base de données:**
les fichiers DICO1 à DICO7 sont annulés, de même que l'INDEX. La base de données est donc vidée et prête à recevoir de nouveaux fichiers.
2. **Génération d'un fichier:**
GENDICO demande le nom du fichier texte à générer et ajoute les informations contenues dans celui-ci aux différents fichiers DICO tout en ajustant l'INDEX.
3. **Gestion de la liste des fichiers:**
GENDICO permet de gérer la liste des fichiers textes à introduire dans la base de données. L'opérateur a la possibilité:
 - d'ajouter un fichier à la liste
 - de supprimer un fichier de la liste
 - de consulter à l'écran la liste des fichiers
 - d'imprimer cette liste
4. **Génération de tous les fichiers:**
tous les fichiers faisant partie de la liste décrite ci-dessus sont ajoutés à la base de données et l'INDEX est ajusté.
5. **Consultation de la liste des mots:**
la liste alphabétique des mots inclus dans la base de données est affichée à l'écran dans un but de consultation.
6. **Impression de la base de données:**
l'opérateur a la possibilité de demander différents types d'impression relatifs aux mots inclus dans la base de données:

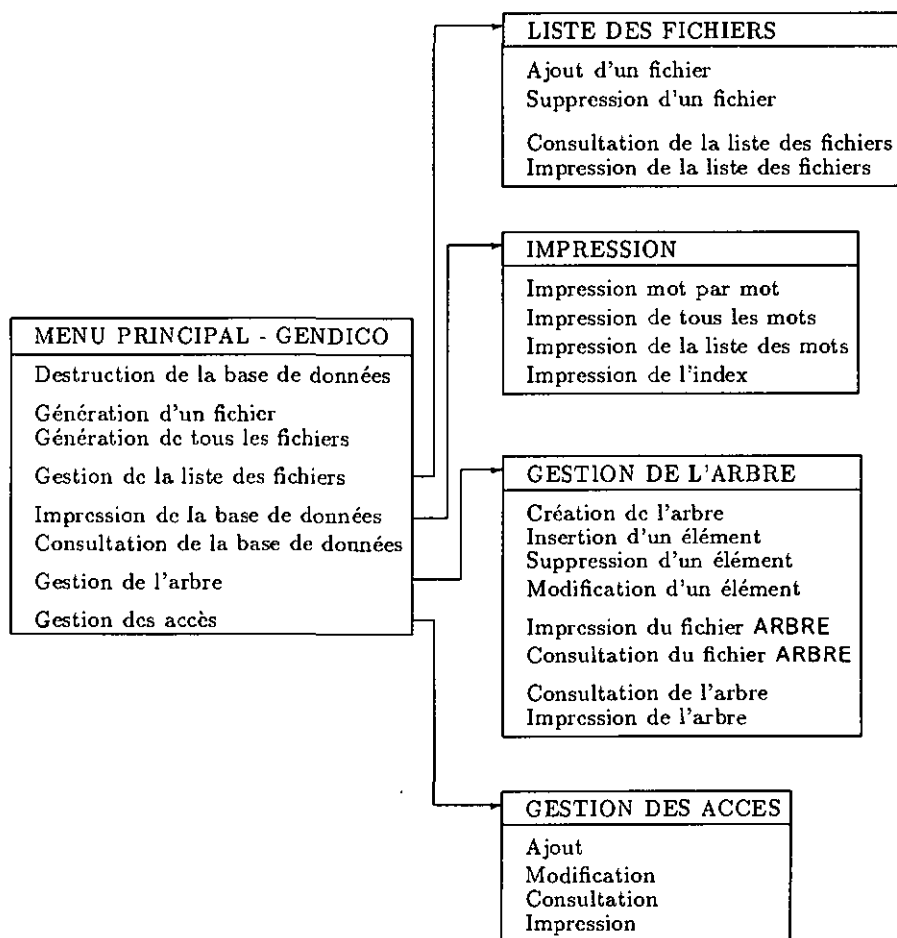


Figure 6.7: Menus de l'interface utilisateur du dictionnaire électronique

- **impression mot par mot:**
impression, selon un cadre de présentation prédéfini, de l'ensemble du texte relatif aux différentes rubriques d'un ou de plusieurs mots choisis par l'opérateur.
- **impression de tous les mots:**
impression de l'ensemble du texte relatif aux différentes rubriques de tous les mots inclus dans la base de données.
- **impression de la liste alphabétique des mots iuclus dans la base de données, avec pour chaque mot une indication concernant les rubriques existant pour chaque mot.**
- **impression de l'index alphabétique des mots.**

7. Gestion de l'arbre:

GENDICO offre l'ensemble des opérations nécessaires à la construction et à la gestion du fichier ARBRE:

- **Création de l'arbre:**
l'ARBRE est initialisé et prêt à être construit. Un nom physique de fichier lui est assigné, de même que le nom de la racine.
- **Insertion d'un élément:**
GENDICO demande le nom du père, et si celui-ci est un élément de la structure, permet de saisir la liste des éléments à lui attribuer comme fils (si le père a déjà des fils, les nouveaux viendront s'ajouter aux précédents).
- **Suppression d'un élément:**
si l'élément demandé existe, il est retiré de l'arbre, de même que l'ensemble de la structure qui lui est rattachée si celui-ci n'est pas un élément terminal de l'arbre.
- **Modification du nom d'un élément**
- **Impression du fichier ARBRE:**
Les champs relatifs à chaque enregistrement du fichier sont imprimés séquentiellement.
- **Consultation du fichier ARBRE:**
Les champs relatifs à chaque enregistrement du fichier sont affichés séquentiellement à l'écran dans un but de consultation.
- **Impression de l'arbre:**
la structure de l'arbre est imprimée selon un parcours en profondeur.

- **Consultation de l'arbre:**

la structure de l'arbre est affichée à l'écran selon un parcours en profondeur dans un but de consultation.

8. Gestion des accès:

le fichier ACCES est géré par l'opérateur au travers des opérations suivantes:

- **Ajout:**

le programme ajuste le fichier ACCES en fonction des mots contenus dans l'INDEX. Pour chaque mot nouveau, il permet de saisir interactivement son équivalent anglais de même que le numéro de logiciel correspondant au mot.

- **Modification:**

l'opérateur peut par cette opération modifier le nom anglais et/ou le numéro de logiciel relatif à un mot spécifique.

- **Consultation:**

les champs relatifs à chaque enregistrement du fichier sont affichés à l'écran dans un but de consultation, par ordre alphabétique des mots français.

- **Impression:**

les champs relatifs à chaque enregistrement du fichier sont imprimés par ordre alphabétique des mots français.

6.4 Méthodes d'accès au dictionnaire électronique

Comme nous l'avons vu, la base de données du dictionnaire électronique est constituée, en plus des fichiers de base DICO1 à DICO7, de trois fichiers particuliers: INDEX, ACCES et ARBRE. C'est par l'intermédiaire de ces trois fichiers que le programme INTDICO (programme d'INTerrogation du DICTiOnnaire) a la possibilité d'accéder aux informations contenues dans les fichiers de base. Ce processus d'accès aux informations est représenté schématiquement en figure 6.8. Nous constatons sur cette figure que le programme INTDICO conduit également aux outils statistiques qui forment le troisième module de l'application globale.

Nous ne nous intéresserons toutefois pas ici à la procédure d'accès aux outils statistiques, celle-ci étant exposée dans le chapitre 8.

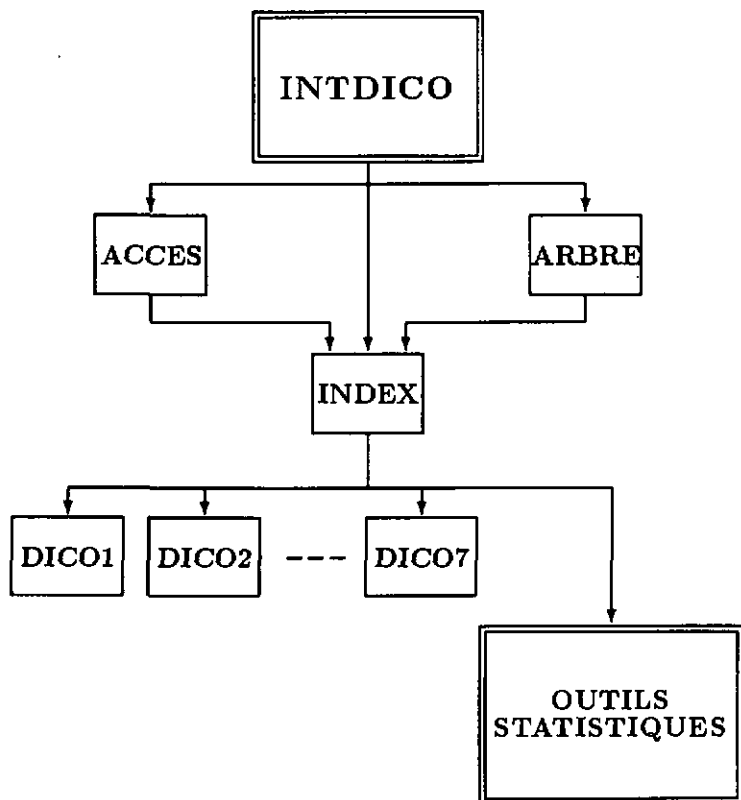


Figure 6.8: Processus d'interrogation du dictionnaire électronique

L'objectif de cette section est de présenter les différentes possibilités d'accès aux fichiers de base par l'intermédiaire des trois fichiers précités.

Les deux cas fondamentaux qui peuvent survenir sont d'une part un accès direct à un mot (lorsque le mot fourni par l'utilisateur est complet), et d'autre part un accès indirect ou un choix dans une liste. Ce second cas apparaît lorsque l'utilisateur fait appel à l'index hiérarchique, ou lorsqu'il fournit un mot incomplet. Le système lui procure alors une liste alphabétique de mots pouvant correspondre à sa demande.

Le tableau 6.3 donne un résumé de ces différentes possibilités.

Type d'accès	Fichier	Algorithme de recherche
Accès direct à une rubrique relative à un mot: <ul style="list-style-type: none"> • par un mot français • par un mot anglais 	INDEX ACCES	recherche dichotomique recherche dichotomique
Accès indirect (choix d'un mot dans une liste): <ul style="list-style-type: none"> • par l'index alphabétique (complet ou partiel) • par l'index hiérarchique 	INDEX ARBRE	recherche de sous-chaîne parcours d'arbre

Table 6.3: Méthodes d'accès au dictionnaire électronique

Nous pouvons constater que ces différents types d'accès nécessitent trois algorithmes de recherche spécifiques, que nous allons développer dans les sections suivantes.

6.4.1 Recherche dichotomique

Le fichier INDEX, de même que le fichier ACCES sont triés par ordre alphabétique des champs respectifs `Index.Enre.Mot` et `Acces.Enre.Mot.Anglais`.

Ce tri permet d'effectuer une recherche dichotomique dans le fichier afin d'optimiser le temps d'accès aux différents enregistrements. Cette recherche fournit en un minimum de pas possibles un accès à l'enregistrement (ou au mot) cherché.

L'idée de la recherche dichotomique est la suivante:

On commence par comparer le mot cherché avec le mot situé au milieu du fichier. Si le résultat de cette comparaison est négatif (le mot du milieu n'est pas identique au mot cherché), ce test nous permettra de déterminer dans quelle moitié du fichier nous devons continuer notre recherche:

- si le mot du milieu est plus petit que le mot cherché, la recherche continue dans la seconde moitié du fichier;
- si, en revanche, le mot du milieu est plus grand que le mot cherché, on continue la recherche dans la première moitié du fichier.

Nous pouvons alors appliquer la même procédure à cette moitié, jusqu'à ce que le résultat de la comparaison soit positif.

Le nombre de pas nécessaires pour trouver le mot cherché ou établir de façon définitive que ce mot n'est pas présent dans le fichier consiste à effectuer au maximum $\text{Trunc}(\log_2(n)) + 1$ comparaisons, où n est le nombre d'enregistrements dans le fichier.

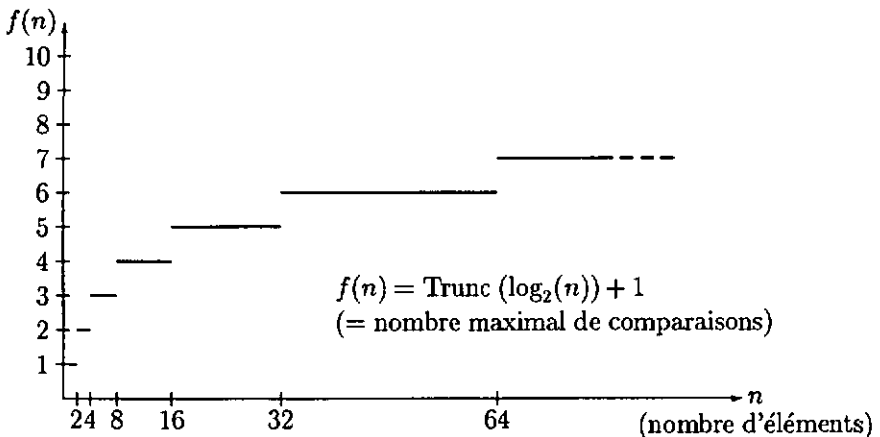


Figure 6.9: Temps d'accès par une recherche dichotomique

*La fonction *Trunc* est une fonction de Turbo Pascal [61] qui renvoie la partie entière d'un nombre réel

La figure 6.9 représente le nombre maximal de comparaisons à effectuer en fonction du nombre n d'enregistrements dans le fichier. Elle montre bien que ce type de recherche est d'autant plus efficace que le nombre d'enregistrements est grand, puisqu'à chaque doublement du nombre d'enregistrements, le nombre maximal de comparaisons n'augmente que de 1.

6.4.2 Recherche de sous-chaînes

Cette seconde méthode d'accès ne fournit pas un accès direct au mot cherché, mais correspond à une situation où l'utilisateur ne connaît qu'une partie du mot cherché. L'objectif de cette méthode est donc de fournir une liste partielle de mots dans laquelle l'utilisateur devra encore effectuer un choix.

L'algorithme que nous avons appelé "recherche de sous-chaînes" peut être considéré de plusieurs manières, car il existe de nombreuses méthodes de comparaison de chaînes de caractères, de recherche d'une chaîne ou d'une sous-chaîne dans un mot.

Une des méthodes consiste à ne saisir qu'une partie du mot cherché (le début du mot par exemple), puis à fournir la liste partielle des mots correspondant à la chaîne saisie. Pratiquement, ceci peut se réaliser en demandant de remplacer la partie manquante du mot par un symbole spécial (par exemple une étoile "★"). Si la chaîne saisie est

TEST★,

l'algorithme doit donc fournir tous les mots du fichier INDEX commençant par les quatre lettres TEST. Le principe est donc le suivant:

1. se positionner par une recherche dichotomique sur le premier mot plus grand ou égal à la chaîne saisie;
2. si le mot trouvé ne correspond pas à la chaîne, l'algorithme fournit une liste vide;
3. sinon, ajouter le mot trouvé à la liste, puis lire successivement les mots du fichier et les ajouter à la liste jusqu'à ce que le mot lu ne corresponde plus à la chaîne cherchée.

Cette méthode a notamment l'avantage de la simplicité. Elle permet également d'obtenir des listes partielles de tous les mots commençant par une certaine

lettre donnée, ou la liste complète de tous les mots si la chaîne saisie se réduit à une simple étoile.

Par contre, elle a le désavantage d'obliger l'utilisateur à entrer une partie spécifique du mot (le début), alors que celui-ci ne connaît pas a priori de quelle manière les mots sont libellés. Un exemple très simple est donné par le terme "ALEATOIRE". Si ce terme spécifique n'est pas directement un mot du dictionnaire, l'utilisateur intéressé par cette notion ne saura pas dans un premier temps sous quel(s) mot(s) il pourra trouver de l'information à ce sujet. Avec la méthode proposée ci-dessus, il sera contraint à parcourir la liste complète des mots du dictionnaire afin d'aboutir à l'information voulue.

Nous proposons donc une autre méthode de recherche qui permet de répondre à ce type d'interrogation. L'idée de base de cette méthode est la recherche de la plus longue sous-chaîne commune à deux chaînes de caractères.

Sur la base d'une chaîne saisie, il sera alors nécessaire de parcourir l'ensemble des mots du fichier afin d'effectuer la recherche de la plus longue sous-chaîne commune à la chaîne saisie et au mot considéré.

Notons $U = (u_1, u_2, \dots, u_m)$ la chaîne saisie (de longueur m) et $V = (v_1, v_2, \dots, v_n)$ un des mots du fichier (de longueur n).

On utilise un tableau d'entiers T à deux dimensions ($m \times n$) et initialisé à 0. Notons L la longueur de la plus longue sous-chaîne trouvée. L'algorithme de recherche de la valeur de L est le suivant:

```

L := 0
For i := 1 To m Do
  For j := 1 To n Do
    If  $u_i = v_j$  Then Begin
      If  $(i = 1)$  Or  $(j = 1)$  Then  $T(i, j) := 1$ 
      Else  $T(i, j) := T(i - 1, j - 1) + 1$ 
      If  $T(i, j) > L$  Then  $L := L + 1$ 
    End
  End

```

Le critère d'inclusion du mot dans la liste doit alors être déterminé:

- une possibilité est de choisir un critère fixe (par exemple, $L \geq 5$). Dans ce cas, tous les mots ayant une sous-chaîne d'au minimum 5 caractères commune avec la chaîne saisie seront inclus dans la liste;

- une autre possibilité est de choisir un critère variable en fonction de la longueur m de la chaîne saisie (par exemple, $L \geq (m \text{ Div} 2)$ ou $L \geq m - 1$). Il est également possible d'obtenir la liste de mots comprenant la chaîne complète en imposant $L = m$.

Le problème du choix du critère n'est pas aisé, car un critère trop restrictif ($L = m$) ne permet pas forcément d'obtenir la liste de tous les mots susceptibles d'être intéressants en fonction de la chaîne saisie: si nous considérons par exemple la chaîne "GRAPHIQUE", celle-ci ne permettra pas d'inclure le mot "GRAPHE" dans la liste.

Un critère trop large par contre permettra d'inclure des mots qui n'ont, outre le fait d'avoir une sous-chaîne en commun, aucun rapport avec la chaîne cherchée: en choisissant par exemple le critère $L \geq (m \text{ Div} 2)$ (ce qui donne pour $m = 9$, $L \geq 4$), nous obtiendrons une liste très hétérogène comprenant par exemple:

- "MOYENNE ARITHMETIQUE",
- "ECHANTILLONNAGE PAR GRAPPES",
- "LOI GEOMETRIQUE",
- "CLASSIFICATION HIERARCHIQUE".

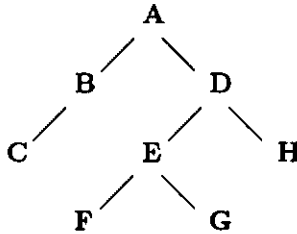
Il s'agit donc de trouver un compromis, en remarquant toutefois qu'un critère plus large, même s'il fournit parfois des informations qui peuvent sembler aberrantes, apporte également dans la grande majorité des cas l'information cherchée; alors qu'un critère plus restrictif, qui ne fournit que rarement des informations aberrantes, peut fréquemment laisser passer des informations qui auraient pu être intéressantes.

6.4.3 Parcours d'arbre

La troisième méthode d'accès que nous proposons est liée à une recherche dans l'index hiérarchique. Celui-ci est stocké dans le fichier ARBRE dont la structure est décrite en section 6.3.3.

Il existe deux méthodes fondamentales de parcours d'un arbre: le parcours en profondeur qui traite consécutivement chacune des branches, et le parcours

en largeur qui traite consécutivement chacun des niveaux. Considérons par exemple l'arbre suivant:



Un parcours préfixé en profondeur donnera la liste:

Liste 1

```

A
  B
    C
  D
    E
      F
      G
    H
  
```

alors qu'un parcours en largeur fournira:

Liste 2

```

A
  B
  D
    C
    E
    H
      F
      G
  
```

Chacun de ces parcours a ses avantages et ses inconvénients. La liste 1 a l'avantage de présenter de manière claire l'ensemble de la structure. Toutefois, il peut être relativement difficile de détecter en un seul coup d'oeil l'ensemble des fils d'un élément, lorsque la structure inférieure de ces fils est complexe (comprend beaucoup d'éléments).

Quant à la liste 2, elle fournit la liste des éléments niveau par niveau. L'ensemble des fils de **A** peut donc être déterminé aisément puisqu'ils sont présentés en bloc. Toutefois, la structure des niveaux inférieurs est perdue puisqu'il est impossible de déterminer par exemple qui est le père de **C**, **E** et **H**, parmi **B** et **D**.

Aucune de ces méthodes ne s'avérant totalement satisfaisante, nous proposons d'opter pour une méthode qui laisse le choix à l'utilisateur de parcourir l'arbre niveau par niveau, à partir d'un nœud qu'il sélectionne lui-même.

Les étapes du processus de parcours de l'arbre sont les suivantes:

1. afficher la racine de l'arbre ainsi que le niveau inférieur composé de l'ensemble de ses fils (ceci équivaut à un parcours en largeur partiel puisque l'on s'arrête à la fin du premier niveau):

```
A
  B
  D
```

puis proposer à l'utilisateur de sélectionner un des éléments affichés.

2. quel que soit l'élément sélectionné, ne conserver parmi les éléments affichés que la liste des pères successifs de cet élément, et lui rajouter le premier niveau de sa structure inférieure (l'ensemble de ses fils).

Si **D** a été sélectionné, nous aurons alors:

```
A
  D
    E
    H
```

Proposer ensuite à l'utilisateur d'effectuer une nouvelle sélection parmi l'ensemble des éléments affichés, et procéder au nouvel affichage en suivant la même règle.

Cette méthode permet ainsi à l'utilisateur d'avoir une vue partielle du parcours en profondeur de l'arbre: la partie du parcours relative à l'élément choisi.

Chapitre 7

Analyse informatique du livre électronique

7.1 Introduction

Le livre électronique est, tout comme le dictionnaire électronique, une application de consultation. Nous exposons dans ce chapitre les points essentiels qui diffèrent dans la construction du livre électronique par rapport à celle du dictionnaire électronique exposée dans le chapitre précédent. Ceci nous amène à développer plus spécifiquement la structure de la base de données relative au livre, et à considérer la démarche du processus de génération de celle-ci. Nous traiterons également quelques aspects spécifiques au processus de consultation du livre électronique.

7.2 Structure de la base de données

La base de données du livre électronique est un peu plus complexe que celle du dictionnaire qui pouvait être représentée par un simple tableau à double entrée.

Celle-ci nécessite notamment la gestion de trois éléments:

1. des chapitres;
2. des sections;
3. des pages.

Ces trois éléments ne sont bien évidemment pas indépendants les uns des autres, mais sont organisés de manière très stricte au sein d'une structure bien définie. Nous pouvons représenter ces éléments par une structure hiérarchique, dont la racine caractérise le livre dans sa globalité.

Le livre est composé d'un certain nombre de chapitres qui constituent l'ensemble de ses fils. A chaque chapitre est alors associée une structure composée de sections et de sous-sections, chaque section étant formée par un ensemble ordonné de pages.

7.2.1 Description des chapitres

Considérons tout d'abord la notion de chapitre et examinons ses caractéristiques:

- un chapitre est l'élément principal du livre: l'ensemble des chapitres forment le livre;
- un chapitre peut être caractérisé par son numéro, et par son titre;
- un chapitre est composé d'un certain nombre de sections. L'accès à un début de chapitre implique donc la recherche de la première section qui le compose;
- le début d'un chapitre peut également être caractérisé par le numéro de la première page de ce chapitre.

L'ensemble de ces caractéristiques sont regroupées dans un fichier que nous appelons `IndCh` (Index relatif aux Chapitres). Chaque enregistrement de ce fichier représente un chapitre. Un enregistrement est donc composé des champs suivants:

```
IndCh.Enre : Record
              No.Chapitre      : Integer
              Titre.Chapitre   : String[60]
              Premiere_Page    : Integer
              Derniere_Page    : Integer
              Premiere_Section : Integer
              Derniere_Section : Integer
End
```

Les champs `No.Chapitre` et `Titre.Chapitre` représentent, comme leur nom l'indique, le numéro et le titre du chapitre. La seule contrainte imposée sur ces champs est relative au numéro du chapitre qui doit obligatoirement être unique. En dehors de cette contrainte, ces deux champs sont suffisamment explicites et ne nécessitent pas de commentaires particuliers.

En revanche, le problème se pose, pour les quatre derniers champs, de déterminer de quelle manière les pages et les sections vont être numérotées. Les réponses à ce problème découlent directement de la description des pages et des sections qui est présentée ci-dessous.

7.2.2 Description des sections

Avant de considérer la structure qui lie les sections entre elles, commençons par examiner les différents éléments caractéristiques d'une section:

- de même que le chapitre, la section peut se caractériser par un numéro, et par son titre;
- une section doit obligatoirement faire partie d'un chapitre. La section peut donc se caractériser par le numéro du chapitre auquel elle appartient;
- en vue d'accéder au contenu d'une section, il est nécessaire de connaître quelles sont les différentes pages de l'ensemble du texte relatives à celle-ci.

Afin de permettre une double numérotation des pages, il est également intéressant de connaître, sur la base d'une numérotation des pages par chapitre, quelles sont les pages relatives à la section considérée;

- nous avons défini en section 4.5.2 qu'une section peut être composée de deux types de pages différents: des pages de texte, et des pages de résumé. Il est donc nécessaire de connaître pour chaque section quel est le nombre de pages de texte, et quel est le nombre de pages de résumé. Une dernière information à connaître est de savoir si la section peut permettre ou non l'accès à un outil statistique, et le cas échéant quel est cet outil.

L'ensemble de ces informations sont stockées dans un fichier appelé **IndSc** (Index relatif aux Sections), dans lequel chaque enregistrement représente une section. Les différents champs relatifs à chaque enregistrement sont les suivants:

```

IndSc.Enre : Record
              No_Section           : Integer
              Titre_Section        : String[60]
              No_Chapitre          : Integer
              Premiere_Page        : Integer
              Derniere_Page        : Integer
              Premiere_Page_Chapitre : Integer
              Derniere_Page_Chapitre : Integer
              Nb_Pages_Texte       : Integer
              Nb_Pages_Resume      : Integer
              Nom_Outil            : String[60]
End
  
```

Nous avons soulevé précédemment le problème de la numérotation des sections. Les sections, dans un livre, ne sont pas organisées séquentiellement, mais de manière hiérarchique. Une section peut donc, selon la terminologie que nous avons présentée au point 6.3.3, avoir des frères et des fils, ces frères et ces fils étant eux-mêmes également des sections.

La numérotation classique d'une telle structure est donnée en figure 7.1.

Cet exemple fictif représente le chapitre I. Le numéro associé à la racine caractérise donc le chapitre. L'ensemble de la structure liée à cette racine représente les différentes section (et sous-sections) y relatives.

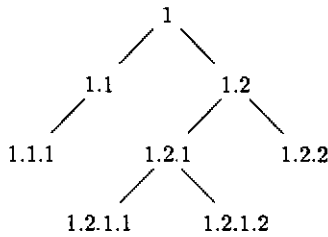


Figure 7.1: Numérotation effective des sections

Une telle numérotation, bien qu'elle ait l'avantage d'être très suggestive, a l'inconvénient de ne pas permettre d'accès direct dans un fichier à l'information relative à un élément particulier. Nous avons donc choisi d'adopter une numérotation séquentielle des sections afin de permettre cet accès direct au fichier.

La numérotation relative au même exemple est donnée en figure 7.2.

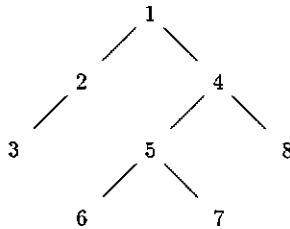


Figure 7.2: Numérotation séquentielle des sections

Nous pouvons définir ici la notion de niveau qui permettra, à partir de cette numérotation séquentielle, de reconstituer le numéro effectif de chaque section: le niveau relatif à un élément est égal au nombre minimal de relations qui permettent de remonter à partir de cet élément jusqu'à la racine de la structure (ici l'élément numéro 1). Ainsi dans cet exemple, les sections 2 et 4 sont de niveau 1, les sections 3, 5 et 8 de niveau 2, et les sections 6 et 7 de niveau 3.

Cette numérotation correspond à une lecture séquentielle du livre. Si l'on poursuit la numérotation des chapitres suivants, la racine du deuxième cha-

pitre se verra donc attribuer le numéro 9. Cette numérotation nous amène à considérer la notion de chapitre au même titre qu'une section, d'un type particulier puisqu'aucune page de texte ne lui est directement associée, mais faisant intégralement partie de la même structure.

Cette structure est conservée séparément dans un fichier nommé `IndSt` (Index relatif à la Structure). Le nombre d'enregistrements dans ce fichier est strictement identique à celui du fichier `IndSc`, chaque enregistrement représentant les éléments nécessaires au positionnement d'une section particulière dans la structure hiérarchique globale du livre. Chaque enregistrement est composé des champs suivants:

```

IndSt_Enre : Record
              No_Element : Integer
              Niveau      : Integer
              Premier_Fils : Integer
              Frere       : Integer
              Soeur       : Integer
              Pere        : Integer
              End

```

Comparativement au fichier `ARBRE` du dictionnaire électronique que nous avons décrit au point 6.3.3, nous avons ajouté ici un champ que nous avons appelé `Soeur`. Comme le montre la figure 7.3, celui-ci représente le "frère de droite" de l'élément, alors que le champ `Frere` représente son "frère de gauche".

Cette information supplémentaire, qui n'est pas indispensable dans la représentation de la structure hiérarchique, a été introduite afin d'accroître la rapidité de la consultation du livre. En effet, la consultation du livre devant permettre un parcours dans le sens habituel ou à rebours, la connaissance du frère de droite et de gauche permet alors un positionnement plus rapide dans les deux sens de lecture.

7.2.3 Description des pages

Le troisième élément à gérer est la notion de page. Une page est définie comme un ensemble de lignes pouvant être affichées simultanément à l'écran. Le nombre maximal de lignes par page a été fixé à 19, ce qui permet de remplir la zone d'information de la page-écran standard présentée au point 4.2.

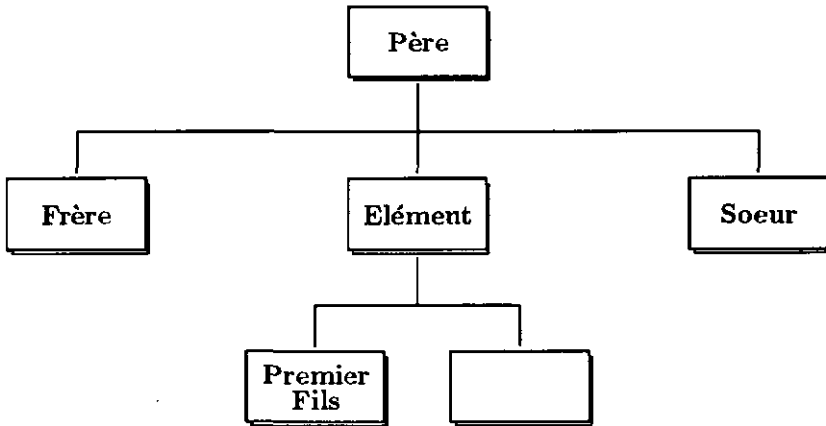


Figure 7.3: Terminologie associée à une structure hiérarchique (2)

La structure sous-jacente à l'ensemble des pages du livre est une structure séquentielle, structure selon laquelle les différentes pages sont numérotées.

Les caractéristiques d'une page sont les suivantes:

- une page pouvant être de deux types (texte ou résumé), la page doit être caractérisée par un statut indiquant quel est le type de la page considérée;
- en plus du numéro de la page (par rapport à la totalité des pages du livre), il est intéressant de connaître le numéro de la page à l'intérieur du chapitre;
- une page est caractérisée d'une part par le numéro du chapitre auquel elle appartient, et d'autre part par le numéro de la section;
- l'affichage d'une page implique la connaissance des lignes de texte relatives à cette page.

Ces différentes caractéristiques sont stockées dans un fichier appelé *IndPg* (Index relatif aux Pages), où chaque enregistrement représente une page particulière. Un enregistrement est donc composé des champs suivants:

```
IndPg_Enre : Record
              No_Page           : Integer
              Status            : Integer
              No_Page.Chapitre  : Integer
              No_Chapitre      : Integer
              No_Section       : Integer
              No_Ligne_Debut   : Integer
              No_Ligne_Fin     : Integer
End
```

Les deux derniers champs `No_Ligne_Debut` et `No_Ligne_Fin` déterminent quelles sont les lignes de texte relatives à la page considérée. Nous avons jusqu'à présent décrit les quatre fichiers "index" nécessaires à la gestion de la structure du livre. Il nous reste maintenant à déterminer comment l'information textuelle doit être stockée, afin d'établir de quelle manière les lignes doivent être numérotées.

7.2.4 Description du texte

Le texte du livre est constitué d'un ensemble ordonné de lignes. Nous avons choisi de conserver l'ensemble des lignes relatives à chaque chapitre dans un fichier particulier. A chaque chapitre est donc associé un fichier. Un chapitre étant caractérisé par un numéro unique, nous pouvons appeler `CHAPx` le fichier relatif au `x`-ème chapitre du livre.

L'accès aux fichiers `CHAPx` se fait obligatoirement par l'intermédiaire de l'index relatif aux pages (`IndPg`) puisque c'est le seul fichier qui contienne les adresses relatives aux numéros de lignes. Les champs `IndPg_Enre.No_Ligne_Debut` et `IndPg_Enre.No_Ligne_Fin` correspondent donc aux numéros de lignes de début et de fin de texte pour chaque page dans le fichier `CHAPx`, où la valeur de `x` est donnée par `IndPg_Enre.No_Chapitre`.

7.3 Description du programme de génération de la base de données

Le programme `GENLIVRE` (programme de GENération du LIVRE électronique) permet de créer et de gérer l'ensemble de la structure de la base de

données telle que nous l'avons décrite.

Il permet également de gérer des informations complémentaires nécessaires à la consultation du livre.

Le processus de génération de la base de données du livre électronique est très semblable à celui du dictionnaire. Il est représenté en figure 7.4.

La tâche de **GENLIVRE** est double:

1. **Générer les fichiers de base du livre:**

Sur la base de fichiers textes saisis par l'opérateur à l'aide d'un éditeur, **GENLIVRE** doit créer automatiquement les différents fichiers de la base de données que nous avons décrits précédemment:

- les quatre fichiers qui conservent la structure du livre: **IndCh** (index relatif aux chapitres), **IndSc** (index relatif aux sections), **IndSt** (index relatif à la structure du livre) et **IndPg** (index relatif aux pages);
- les fichiers (**CHAP1** à **CHAPn**) qui contiennent l'ensemble du texte du livre.

2. **Gérer l'index du livre:**

Deux fichiers supplémentaires sont également créés par **GENLIVRE**: les fichiers **MOT** et **IND**, qui conservent l'index du livre dont nous avons parlé au point 4.5.3.

7.3.1 Génération des fichiers de base du livre

Tout comme pour le dictionnaire électronique, les fichiers texte doivent avoir été saisis selon des critères de formatage spécifiques afin de permettre à **GENLIVRE** de générer la base de données.

Le premier critère est qu'il doit exister un et un seul fichier par chapitre. Chaque fichier doit ensuite être structuré de telle manière que **GENLIVRE** puisse repérer:

- le numéro et le titre du chapitre;
- le début, la fin, le numéro et le titre de chaque section;

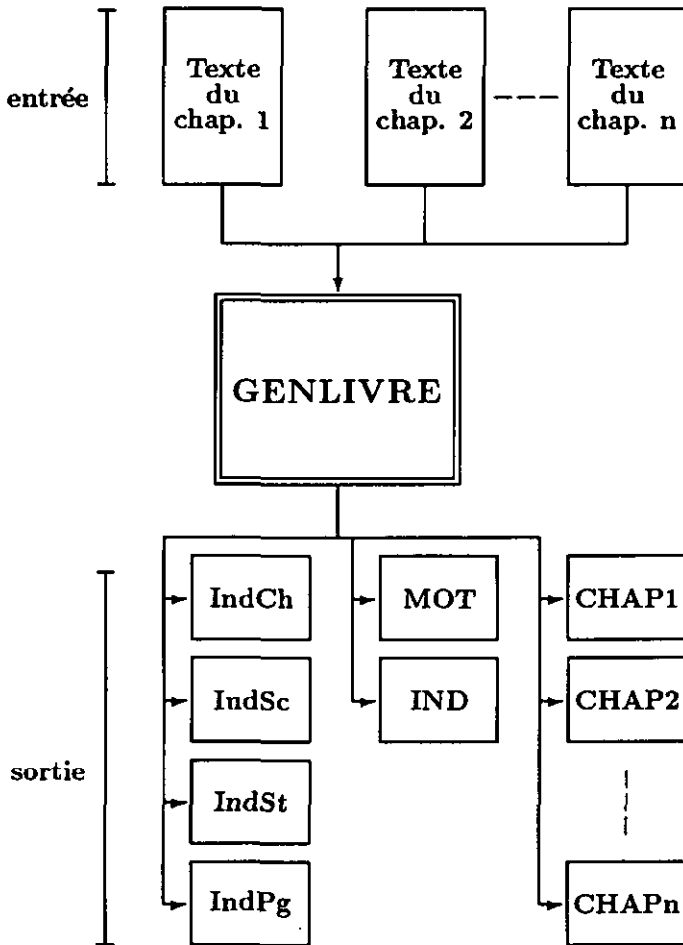


Figure 7.4: Processus de génération du livre électronique

- le début et la fin de chaque page, ainsi que le type de page (texte ou résumé).

Une cellule sera entrée selon le schéma de la figure 7.5:

première ligne	:	..chapitre X Titre_Chapitre (X = numéro du chapitre)
début d'une section	:	..section Titre.Section
début d'une sous-section	:	..ssection Titre.Section le nombre de 's' indique le niveau de la section, ce qui permettra de reconstituer son numéro.
début d'une page résumé	:	..resume
début d'une nouvelle page	:	..page
accès à un outil statistique	:	..outil Nom.Outil le nom de l'outil doit correspondre à un mot du dictionnaire.
fin du chapitre	:	..fin

Les pages de texte doivent obligatoirement être entrées avant les pages résumé à l'intérieur d'une même section.

Certaines règles supplémentaires sont identiques aux règles que nous avons déjà décrites en section 6.3.1 pour la saisie des textes du dictionnaire. Celles-ci sont relatives d'une part à l'inclusion d'images graphiques à l'intérieur des textes, et d'autre part à la mise en évidence des mots accessibles par le dictionnaire électronique.

7.3.2 Gestion de l'index du livre

L'index doit offrir la possibilité d'accéder directement à une page spécifique en fonction d'un mot donné par l'utilisateur (mot qu'il choisit dans une liste).

La gestion de l'index se fait par deux fichiers:

- MOT qui contient la liste des mots de l'index. Pour chaque mot, il conserve un pointeur qui permet l'accès au fichier IND:

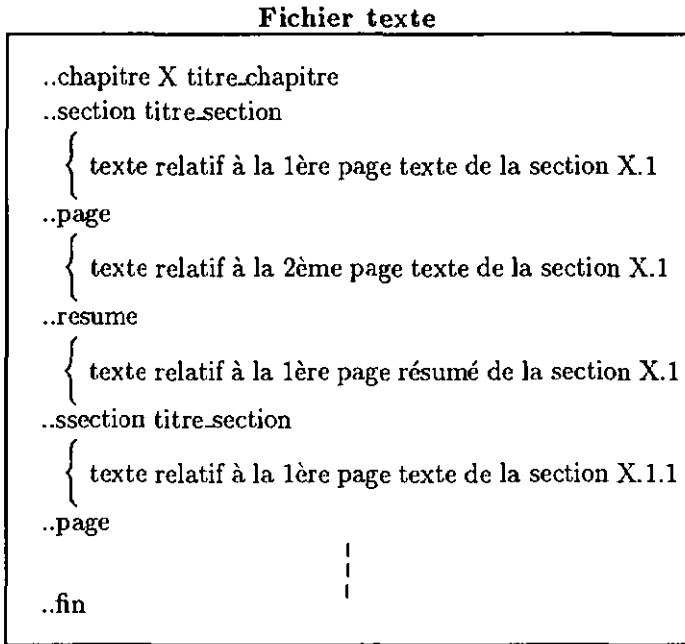


Figure 7.5: Structure d'un fichier texte (input) du livre électronique

```

Mot_Enre : Record
           Mot      : String[60]
           Pointeur : Integer
           End

```

Le champ `Mot_Enre.Pointeur` est l'adresse du premier enregistrement du fichier IND relatif au mot considéré.

- IND qui contient pour chaque mot un ensemble d'enregistrements contenant les informations relatives aux pages accessibles pour chaque mot: le numéro du chapitre, le numéro de la page, de même que le numéro de la page par rapport au chapitre considéré. La description d'un tel enregistrement est la suivante:

```

Ind_Enre : Record
           Mot      : String[60]
           No_Chapitre : Integer
           No_Page   : Integer
           No_Page_Chapitre : Integer
           End

```

La figure 7.6 représente schématiquement l'accès au fichier IND par le fichier MOT.

La construction de l'index se fait en deux étapes:

1. la première étape consiste à créer la liste des mots devant figurer dans l'index (fichier MOT).
2. la seconde étape consiste à associer à chaque mot de la liste les références aux pages du livre électronique. Cette opération peut être conçue de différentes manières:
 - (a) **automatique**: le programme recherche tous les numéros de page pour chaque mot de la liste. La tâche de l'opérateur consiste donc uniquement à appeler la procédure, et le travail d'association des numéros de page aux mots se réalise automatiquement. Une telle procédure possède toutefois quelques inconvénients: d'une part, toutes les pages auxquelles le mot a été rencontré seront incluses dans l'index, alors que certaines d'entre elles ne le mériteraient peut-être pas. D'autre part, certaines pages mériteraient peut-être

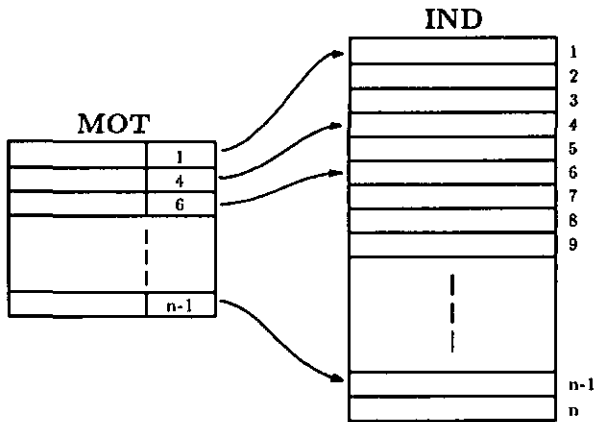


Figure 7.6: Représentation de l'index du livre électronique

de figurer dans l'index alors que le mot ne figure pas explicitement sur la page.

- (b) **assistée par mot**: le programme affiche chaque page du livre et propose de garder ou de supprimer cette page comme référence pour le mot traité.

Dans une telle procédure, l'ensemble du texte doit être parcouru pour chaque mot.

- (c) **assistée par page**: le programme affiche chaque page et demande à l'opérateur d'entrer la liste des mots pour lesquels cette page doit être donnée en référence.

L'ensemble du texte n'est alors parcouru qu'une seule fois, mais le travail de l'opérateur est plus complexe dans le sens où il doit considérer à chaque page tous les mots de la liste.

- (d) **manuelle**: le travail de l'opérateur consiste à entrer manuellement pour chaque mot la liste des pages correspondantes.

Une telle procédure demande un grand travail préalable de l'opérateur qui doit constituer l'index sur papier sans aide du programme avant de l'entrer sur ordinateur.

Nous proposons une combinaison de ces possibilités qui consiste dans un premier temps à créer l'index de manière automatique, puis offre la possibilité d'ajouter manuellement de nouvelles références, ou d'en

supprimer par un contrôle assisté du travail réalisé par la procédure automatique.

7.3.3 Interface opérateur

De même que pour le dictionnaire électronique, la première tâche de l'opérateur est la saisie des textes de base en fonction des critères définis ci-dessus. **GENLIVRE** met ensuite à disposition une interface de travail lui permettant de créer et de gérer la base de données. Cette interface se présente sous forme de menus dont la structure est donnée en figure 7.7.

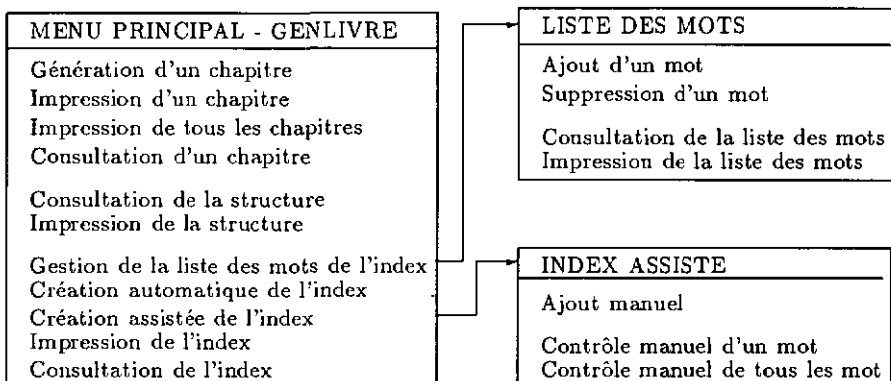


Figure 7.7: Menus de l'interface opérateur du livre électronique

Les possibilités offertes par **GENLIVRE** sont les suivantes:

1. Génération d'un chapitre:

la procédure de génération d'un chapitre est totalement automatique. Le seul travail de l'opérateur est d'entrer le nom du fichier contenant le texte du chapitre à inclure dans la base de données. Si ce chapitre existe déjà dans la base de données, l'ancien chapitre est remplacé par le nouveau. Les fichiers conservant la structure du livre sont ajustés:

- le chapitre est ajouté (ou remplacé) dans `IndCh`;
- les fichiers de gestion des sections `IndSc` et `IndSt` sont mis à jour afin que les numéros de sections soient toujours attribués selon l'ordre logique de parcours (lecture) des sections;

- le fichier de gestion des pages est mis à jour afin que les numéros de pages soient également attribués selon l'ordre logique de parcours des pages, les pages textes précédant les pages résumés à l'intérieur d'une section.

2. Impression d'un chapitre:

cette fonction d'impression fournit une sortie sur papier de l'ensemble du texte relatif à un chapitre. La structure des pages relative au livre électronique n'est pas prise en compte. Les pages sont donc restructurées et renumérotées: à la première page du chapitre est attribué le numéro 1, et les pages suivantes sont numérotées séquentiellement, la longueur d'une page étant de 60 lignes.

3. Impression de tous les chapitres:

l'impression de tous les chapitres fournit une sortie sur papier de l'ensemble du texte inclus dans la base de données du livre électronique.

4. Consultation d'un chapitre:

la fonction de consultation d'un chapitre donne un accès direct au programme de consultation du livre électronique.

5. Consultation de la structure du livre:

le contenu des fichiers de gestion de la structure du livre sont affichés à l'écran. La consultation de ces fichiers permet à l'opérateur de vérifier si les chapitres ont été générés correctement.

6. Impression de la structure du livre:

le contenu des fichiers de gestion de la structure du livre sont imprimés.

7. Gestion de la liste des mots de l'index:

le fichier MOT est géré par l'opérateur au travers des opérations suivantes:

- l'ajout d'un mot dans le fichier. Celui-ci est trié par ordre alphabétique en fin d'ajout;
- la suppression d'un mot dans le fichier. Le fichier IND, de même que les pointeurs de MOT sont mis à jour en fin de suppression;
- la consultation à l'écran de la liste des mots;
- l'impression de la liste des mots.

8. Création automatique de l'index:

la création automatique de l'index est une recherche dans tous les chapitres générés de tous les mots de la liste du fichier MOT.

Le fichier IND est créé (ou remplacé si celui-ci existait déjà). A chaque mot de la liste sont associés tous les numéros de page auxquels celui-ci a été trouvé.

Cette procédure est relativement longue puisqu'elle implique une comparaison successive de tous les mots de la liste avec toutes les lignes de texte du livre.

9. Création assistée de l'index:

le fichier IND créé automatiquement par la fonction précédente peut être modifié manuellement:

- par un ajout manuel: l'opérateur qui veut ajouter une entrée dans le fichier IND doit indiquer le mot et le numéro de page correspondant. Le mot voulu doit exister dans le fichier MOT pour que l'ajout soit valable.

Cette fonction peut être utile pour ajouter dans l'index des références à des pages où le mot n'apparaît pas explicitement, mais qui peuvent aider à sa compréhension.

- par un contrôle assisté d'un mot: **GENLIVRE** affiche successivement le contenu des pages pointées par IND pour le mot voulu. Après l'affichage de chaque page, il est alors possible de confirmer la présence de la page dans l'index ou de la supprimer.
- contrôle assisté de tous les mots: **GENLIVRE** parcourt la liste complète des mots du fichier MOT, et il procède à l'affichage successif du contenu des pages pointées par IND pour chaque mot. Cette fonction permet une vérification complète de l'état de l'index.

10. Impression de l'index:

la liste complète des mots de l'index ainsi que leurs numéros de page associés sont imprimés.

11. Consultation de l'index:

la même liste est affichée à l'écran dans un but de consultation.

7.4 Consultation du livre électronique

Le programme **CONSLIVRE** (programme de CONSultation du LIVRE électronique) constitue le deuxième module de l'application globale de Formation Assistée par Ordinateur. Celui-ci met à disposition les différentes fonctionnalités que nous avons décrites en section 4.5 et qui permettent de consulter la base de données du livre.

Le processus de consultation du livre est représenté schématiquement en figure 7.8.

Nous pouvons constater sur ce schéma qu'il existe deux embranchements différents menant à la consultation des fichiers CHAP1 à CHAPn:

1. l'embranchement de gauche (par MOT et IND) correspond à une recherche par l'index,
2. et l'embranchement de droite (par IndCh et IndSc) correspond à une recherche par la table des matières.

Le tableau suivant présente les différents processus de recherche d'une page spécifique au travers de chacun des deux embranchements:

Embranchement	Processus de recherche d'une page spécifique
recherche par l'index	<ul style="list-style-type: none"> • repérer le mot choisi dans le fichier MOT par une recherche dichotomique dans le fichier; • trouver le numéro de page spécifique par un accès direct au fichier IND grâce au champ <i>Pointeur</i> du fichier MOT.
recherche par la table des matières	trouver le numéro de la première page: <ul style="list-style-type: none"> • du chapitre par un accès direct au fichier <i>IndCh</i> si le numéro du chapitre choisi est connu; • de la section par un accès direct au fichier <i>IndSc</i> si le numéro de la section choisie est connu.

Que l'on choisisse l'un ou l'autre des embranchements, l'accès au texte se fait

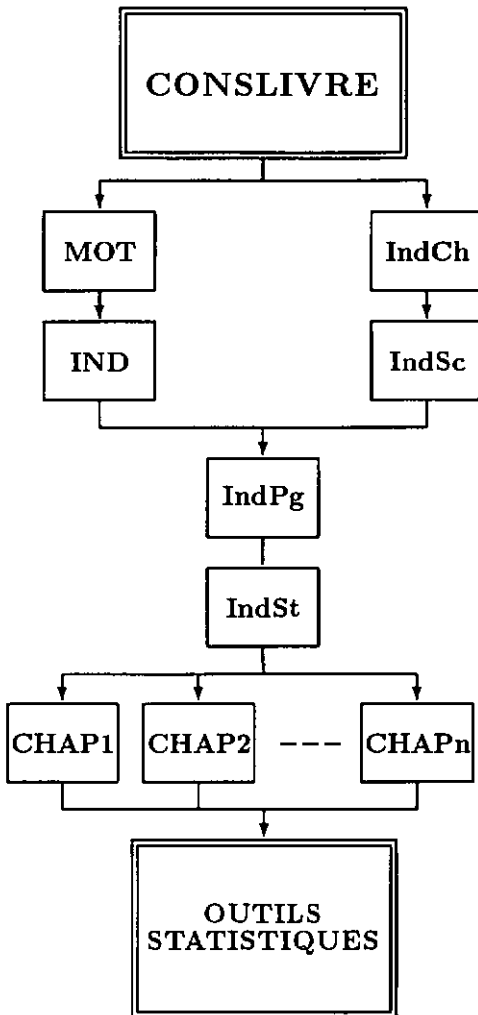


Figure 7.8: Processus de consultation du livre électronique

par un accès direct au fichier `IndPg` qui fournit les numéros de lignes du texte à afficher.

En dehors de la recherche dichotomique que nous avons exposée en section 6.4.1 dans le cadre de l'analyse du dictionnaire électronique, les différents processus de recherche d'une page particulière lors de la consultation du livre électronique sont essentiellement basés sur des méthodes d'accès direct aux fichiers.

Nous ne détaillerons pas ces différents processus, ceux-ci étant d'une complexité relativement faible. Par contre, nous allons approfondir un problème tout-à-fait spécifique à la consultation d'un livre, qui est celui de la reconstitution du numéro effectif d'une section (numéro à afficher), étant donné son numéro séquentiel (numéro associé à chaque section en fonction de l'ordre logique de lecture). Nous pouvons constater sur le schéma représenté en figure 7.8 que, quel que soit l'embranchement choisi, l'accès au texte du livre nécessite l'utilisation du fichier `IndSt` qui contient la structure des sections. C'est en effet grâce aux informations contenues dans ce fichier qu'il est possible de reconstituer le numéro effectif d'une section.

La nécessité de reconstituer le numéro effectif d'une section peut se produire dans trois cas différents:

1. lors du positionnement à un endroit quelconque du livre (par exemple lors de l'accès à une page particulière au travers de l'index);
2. lors du passage à la section suivante en cas de lecture séquentielle descendante des pages du livre (parcours dans le sens habituel de lecture);
3. lors du passage à la section précédente en cas de lecture séquentielle ascendante (parcours à rebours).

Dans le premier cas, le numéro effectif de la nouvelle section doit être entièrement construit, alors que dans les deux derniers cas, il peut être reconstitué à partir du numéro effectif de la section en cours de consultation.

Examinons les trois cas que nous avons mentionnés ci-dessus.

1. Cas 1: positionnement

Considérons la structure fictive donnée par les figures 7.1 et 7.2 ci-dessus. Le problème est de reconstituer le numéro effectif donné en figure 7.1, à partir de la numérotation séquentielle de la figure 7.2.

Les données que nous utilisons sont les suivantes:

numéro séquentiel	niveau	frère	père	numéro effectif à reconstituer
1	0	0	0	1
2	1	0	1	1.1
3	2	0	2	1.1.1
4	1	2	1	1.2
5	2	0	4	1.2.1
6	3	0	5	1.2.1.1
7	3	6	5	1.2.1.2
8	2	5	4	1.2.2

Notons le numéro effectif d'une section par:

$$X_1.X_2.\dots.X_{n-1}.X_n,$$

où les X_i ($i = 1, \dots, n$) représentent les différentes composantes numériques du numéro effectif, n étant le nombre de composantes (qui correspond au niveau de la section incrémenté de 1).

L'algorithme que nous présentons et qui permet de reconstituer ce numéro effectif utilise la terminologie suivante:

- **No_Section**: numéro séquentiel de la section pour laquelle nous voulons reconstituer le numéro effectif
- **No_Chapitre**: numéro du chapitre qui contient la section considérée.
- **Niveau(X)**: niveau de la section X
- **Frere(X)**: numéro du frère (de gauche) de la section X
- **Pere(X)**: numéro du père de la section X

On utilise de plus deux variables entières i et **Element**.

L'algorithme est alors le suivant:

- (a) Initialisations:
 Element := No.Section
 i := Niveau(Element) + 1
- (b) Effectuer (c) jusqu'à ce que Element = 0
- (c) Détermination de la valeur de X_i :
 $X_i := 1$
 Tant que Frere(Element) $\neq 0$:
 $X_i := X_i + 1$
 Element := Frere(Element)
 $i := i - 1$
 Element := Pere(Element)
- (d) Détermination de la valeur de X_1 :
 $X_1 := \text{No.Chapitre}$

L'algorithme effectue donc un parcours à rebours: en partant de la section considérée, on parcourt toutes les sections inférieures du même niveau au moyen de la fonction Frere(X). Lorsque toutes les sections inférieures du même niveau ont été parcourues (Frere(Element) = 0), on remonte d'un niveau dans la structure par la fonction Pere(X). L'algorithme s'arrête lorsque tous les niveaux ont été parcourus, c'est-à-dire lorsque Pere(Element) = 0.

2. Cas 2: lecture séquentielle descendante

Dans le cas d'une lecture séquentielle, on connaît le numéro effectif de la section en cours. Le numéro suivant à rechercher n'a donc pas besoin d'être totalement reconstruit, mais peut être reconstitué par une modification du numéro effectif de la section en cours, que nous notons:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-1} \cdot X_n.$$

Nous utilisons pour cela la notion de niveau. Appelons Niveau_Actuel le niveau de la section en cours de consultation, et Niveau_Suivant le niveau de la section suivante.

Lors du passage d'une section à la section suivante, trois cas peuvent se produire:

- (a) Niveau_Suivant = Niveau_Actuel

dans ce cas, il suffit d'incrémenter la valeur de X_n de 1 pour obtenir le numéro effectif de la section suivante. On obtient alors:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-1} \cdot (X_n + 1).$$

- (b) Niveau_Suivant > Niveau_Actuel

dans ce cas, il faut ajouter l'élément .1 à la chaîne qui devient égale à:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-1} \cdot X_n \cdot 1.$$

- (c) Niveau_Suivant < Niveau_Actuel

si l'on remonte d'un ou plusieurs niveaux, il est alors nécessaire de supprimer de la chaîne les k derniers éléments X_i , où k représente la différence entre Niveau_Actuel et Niveau_Suivant. On obtient alors:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-k-1} \cdot X_{n-k}.$$

Il reste à incrémenter X_{n-k} de la chaîne ainsi obtenue de 1 pour obtenir le numéro effectif de la nouvelle section:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-k-1} \cdot (X_{n-k} + 1).$$

3. Cas 3: lecture séquentielle ascendante

Le cas de la lecture séquentielle ascendante est très semblable à celui de la lecture séquentielle descendante.

Appelons Niveau_Actuel le niveau de la section en cours de consultation, et Niveau_Precedent le niveau de la section précédente.

Lors du passage d'une section à la section précédente, trois cas peuvent également se produire:

- (a) Niveau_Precedent = Niveau_Actuel

dans ce cas, il suffit de décrémenter la valeur de X_n de 1 pour obtenir le numéro effectif de la section suivante:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-1} \cdot (X_n - 1).$$

- (b) Niveau_Precedent < Niveau_Actuel

si l'on remonte d'un ou plusieurs niveaux, il est alors nécessaire de supprimer de la chaîne les k derniers éléments X_i , où k représente la différence entre Niveau_Actuel et Niveau_Precedent. On obtient alors le numéro effectif de la nouvelle section:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-k-1} \cdot X_{n-k}.$$

(c) Niveau_Precedent > Niveau_Actuel

si le niveau de la section précédente est plus grand que le niveau de la section en cours, le problème est un peu plus complexe, car après avoir décrémenté le dernier élément X_n de 1:

$$X_1 \cdot X_2 \cdot \dots \cdot X_{n-1} \cdot (X_n - 1),$$

il est encore nécessaire de reconstruire la partie du chemin parcouru dans la structure hiérarchique pour aboutir à la nouvelle section cherchée. Or, cette reconstitution n'est pas aisée, car les informations dont nous disposons ne nous indiquent pas a priori dans quelle branche de la structure se trouve la section cherchée.

Dans ce cas, il est donc préférable (plus rapide) de recourir à l'algorithme présenté au cas 1 afin de reconstruire le numéro effectif de la nouvelle section.

Chapitre 8

Analyse informatique des outils statistiques

8.1 Introduction

Un outil statistique est une routine permettant de soumettre un ensemble de données à un traitement statistique spécifique.

Chaque outil statistique est rattaché à un mot figurant dans la base de données du dictionnaire électronique.

Un outil statistique est en soi une application interactive de calcul telle que nous l'avons exposée en section 5.3. L'outil doit donc offrir d'une part la possibilité de saisir des informations, et d'autre part celle de recevoir des résultats.

L'interface de saisie doit permettre d'entrer des données et de donner des

ordres de traitement de ces données.

L'interface de présentation des résultats a pour but de mettre les résultats à la disposition de l'utilisateur.

8.2 Accès aux outils statistiques

L'accès à un outil statistique est autorisé par les deux modules de base de l'application globale:

1. par le programme d'interrogation du dictionnaire électronique **INT-DICO**, présenté en figure 6.8, qui permet l'accès à l'outil correspondant au mot en cours de consultation à un moment donné;
2. par le programme de consultation du livre électronique **CONSLIVRE**, présenté en figure 7.8, qui permet d'accéder à l'outil correspondant à la matière en cours de lecture.

Les programmes d'exécution des outils statistiques sont regroupés dans un certain nombre de routines indépendantes.

Ce regroupement est fait notamment en fonction de la similitude de leur procédure de saisie des données, mais également en fonction de la taille des programmes, une routine Turbo Pascal ne pouvant contenir que jusqu'à 64 Ko de code machine.

L'ensemble des routines forment la bibliothèque des routines disponibles.

Le schéma d'appel d'un outil statistique est représenté en figure 8.1. Selon ce schéma, on peut constater que, quel que soit le module appelant, les demandes d'accès à un outil particulier sont récupérées dans un programme de transfert appelé **OUTILS**. La tâche de celui-ci consiste à:

- récupérer le nom de l'outil appelé,
- vérifier l'existence de cet outil dans la bibliothèque de routines, et le cas échéant de transférer le contrôle à la routine d'exécution de l'outil,
- rendre le contrôle au programme appelant en fin d'exécution de l'outil.

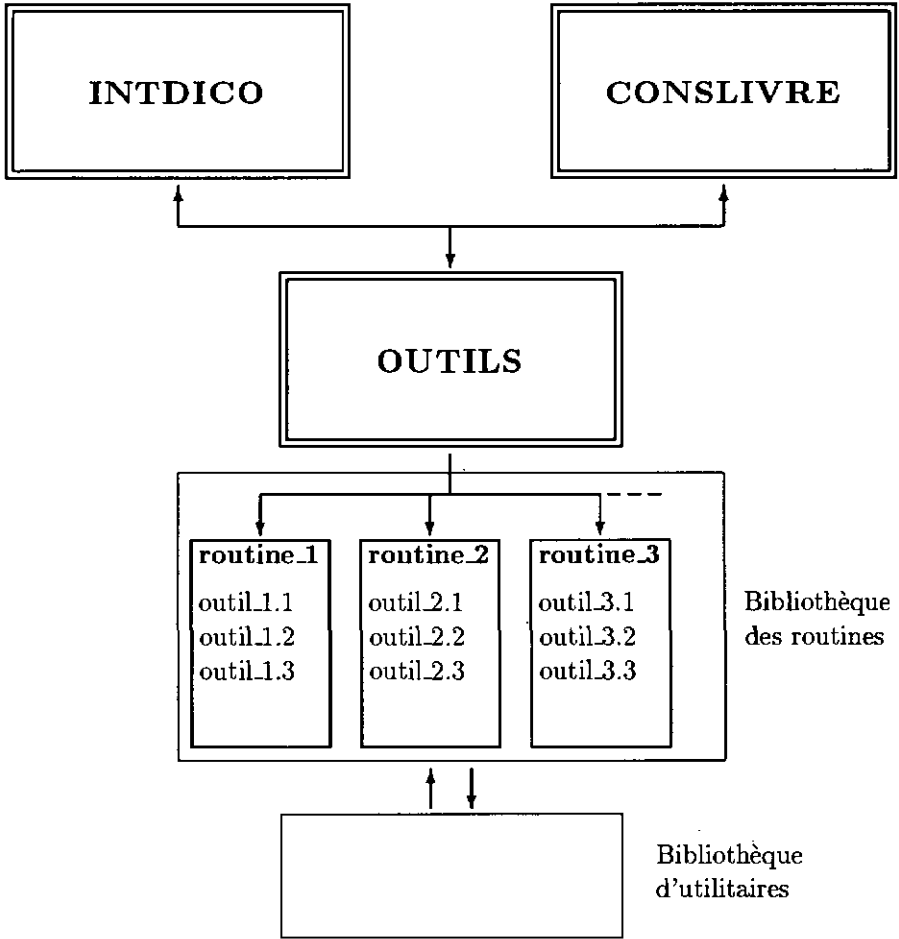


Figure 8.1: Schéma d'appel d'un outil statistique

Les routines sont liées à une bibliothèque d'utilitaires qui met à disposition un ensemble de procédures et de fonctions standard. Une partie de ces utilitaires sont des unités standard de Turbo Pascal:

Crt Contrôle du mode d'affichage, des couleurs, du son, manipulation de fenêtres, accès aux codes étendus des touches du clavier.

Dos Accès à des fonctions du DOS (contrôle de la date et de l'heure système, recherche de répertoires, lancement d'autres programmes).

Graph Module de gestion de l'écran en mode graphique.

D'autres utilitaires ont été développés pour notre application spécifique et permettent d'assurer une uniformité de présentation de l'ensemble des outils:

DefData Définition des types et des variables globales.

Souris Utilitaire de contrôle de la souris.

UtilData Utilitaire de gestion des données (saisie et manipulation de données).

UtilImp Utilitaire d'impression des données.

UtilImpG Utilitaire d'impression de graphiques.

UtilTab Utilitaire d'accès aux tables statistiques.

UtilCalc Utilitaire de calcul.

8.3 Structure d'une routine

Une routine peut permettre l'exécution d'un seul outil, ou de plusieurs outils de structure similaire. A chaque outil est associé un code alphanumérique appelé **Genre**, code qui est transmis à la routine par le programme de transfert **OUTILS**. Chaque outil spécifique à une routine possède donc un **Genre** unique, qui permet d'effectuer les embranchements nécessaires dans le cadre de la routine appelée.

Toutes les routines sont construites selon une structure de base similaire.

La procédure principale effectue une boucle permettant de récupérer le choix de l'utilisateur, puis d'appeler la procédure correspondant à ce choix:

Proc.Donnees Procédure de saisie et manipulation de données.

Proc.Traitement Procédure de traitement des données. Celle-ci constitue la seule phase de calcul de la routine. Elle affiche en sortie le résultat du traitement, ainsi que les éventuels messages d'erreur (par exemple l'insuffisance du nombre de données).

Proc.Formule Procédure d'affichage des formules utilisées dans le traitement.

Proc.Explication Procédure d'affichage d'explications complémentaires au résultat affiché par **Proc.Traitement**.

Proc.Impression Procédure d'impression des données ainsi que de l'ensemble des informations préalablement demandées par l'utilisateur, relatives à ces données.

Proc.Sortie Sortie de l'outil statistique en cours. Le contrôle est rendu au programme **OUTILS**.

Si l'on exclut la procédure **Proc.Sortie** qui a pour seule tâche de quitter l'outil en cours, on peut regrouper ces différentes procédures en deux groupes: des dispositifs d'entrées (procédure de saisie des données: **Proc.Donnees**) et des dispositifs de sorties (procédures d'affichage de résultats: **Proc.Traitement**, **Proc.Formule** et **Proc.Explication**; et procédure d'impression: **Proc.Impression**).

Les deux sections suivantes donnent une description plus détaillée du fonctionnement de chacun de ces dispositifs.

8.3.1 Dispositifs de sortie des résultats

Les dispositifs de présentation des résultats à l'écran constituent un des deux aspects fondamentaux de l'interface de dialogue dans une application interactive de calcul. Leur informatisation ne pose toutefois pas de grands problèmes techniques.

Une fois que les résultats à transmettre ont été définis, il reste à déterminer techniquement comment modéliser ces résultats afin de les afficher sur la place

disponible à l'écran. Comme nous l'avons mentionné en section 4.3.2, la place réservée à l'affichage des résultats correspond à la moitié inférieure de la zone d'information relative à la page-écran standard. Ceci correspond à une fenêtre composée de 10 lignes et 80 colonnes.

Un des moyens les plus simples est de définir cette fenêtre comme cadre déroulant, et de présenter tous les résultats séquentiellement dans ce cadre. Dans ce cas, il n'est pas nécessaire de se préoccuper de la place nécessaire à l'affichage des informations (de la taille d'un tableau, de la longueur d'une formule, etc).

Nous avons toutefois opté pour une solution différente, qui force à une plus grande structuration des résultats. Ceux-ci peuvent être classés en cinq catégories:

1. résultat numérique simple d'un traitement
2. formule de base au calcul du résultat
3. explication courte
4. explication longue (tableaux explicatifs, tests complémentaires, etc)
5. représentation graphique

Les deux derniers types de résultats nécessitent l'utilisation d'écrans complets (en mode texte ou graphique), ce qui ne nous permet pas de les intégrer dans le cadre précédemment défini.

En revanche, les trois premiers types de résultats peuvent relativement facilement entrer dans ce cadre, qui sera donc selon les besoins divisé en deux ou trois parties. Des exemples en sont présentés en figure 8.2.

Les informations à transcrire dans ces différents cadres peuvent être de deux natures:

1. du texte explicatif
2. des résultats numériques

Si l'affichage de texte ne pose pas de problème particulier, il n'en va pas de même pour les résultats numériques. Il est nécessaire de définir certaines règles

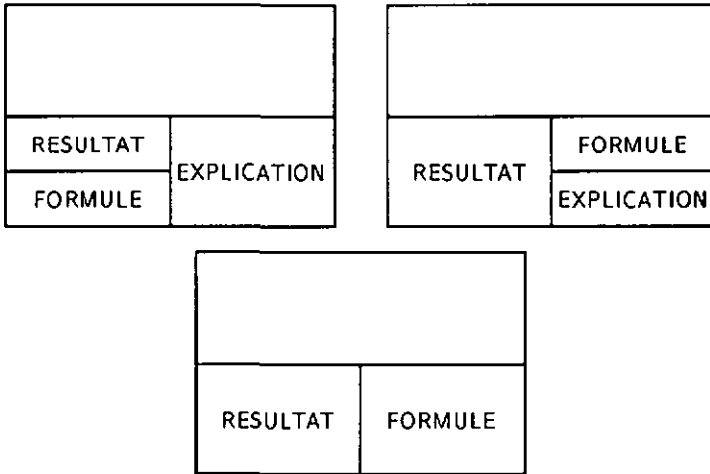


Figure 8.2: Fenêtre de présentation des résultats

d'affichage, permettant de gérer au mieux les très grands nombres (en valeur absolue) ainsi que les très petits nombres (également en valeur absolue).

L'affichage permanent en notation scientifique est une solution possible à ce problème, car il permet de fixer un nombre de caractères maximal d'affichage de n'importe quelle valeur. On notera par exemple 5.00E2 pour 500, -1.20E-1 pour -0.12, 2.48E-8 pour 0.0000000248, ou -6.75E12 pour -6750000000000. Cette notation est tout-à-fait adéquate dans les deux derniers cas, car la lecture de tels nombres n'est pas aisée. Toutefois, elle n'est pas toujours indispensable. Il s'agit donc de fixer à partir de quelles limites (inférieure et supérieure) un nombre doit être affiché en notation scientifique.

La solution que nous proposons se base sur le fait que notre perception visuelle nous permet de lire de manière aisée un nombre relativement petit (par exemple un nombre composé d'un maximum de six chiffres). Nous proposons donc l'utilisation de la notation scientifique, pour les valeurs entières et réelles, uniquement entre les bornes suivantes:

- $] -\infty; -10^6]$
- $[-10^{-6}; 10^{-6}]$
- $[10^6; \infty[$

En dehors de ces bornes, les valeurs réelles sont affichées avec un standard de trois décimales.

8.3.2 Dispositifs d'entrée de données

Après avoir donné quelques indications liées à l'informatisation des dispositifs de présentation des résultats, nous traiterons ici de quelques problèmes liés à l'informatisation des dispositifs d'entrée de données.

La saisie des données est une étape essentielle dans toute application interactive de calcul.

Une solution adoptée dans la plupart des logiciels traditionnels de statistique est de séparer les phases de saisie et de traitement des données. L'utilisateur entre ses données de manière totalement libre (indépendamment du traitement qu'il désire effectuer), puis il indique au système à quoi correspondent les différentes lignes ou colonnes de données entrées. Dans une telle situation, c'est donc à l'utilisateur de déterminer sous quelle forme il doit entrer ses données afin de pouvoir ensuite leur appliquer les traitements voulus.

L'objectif des outils statistiques étant principalement de répondre à un besoin d'apprentissage, nous avons choisi d'opter pour une saisie des données plus dirigée, et de personnaliser l'entrée des données pour chaque outil. Notre but est ainsi de guider l'utilisateur dans la manipulation du logiciel, de même que dans la compréhension des possibilités d'utilisation de chaque méthode statistique.

La saisie des données dans un environnement assisté pose plusieurs problèmes. Nous aborderons ici deux problèmes fondamentaux: 1. le parcours logique de saisie des données; 2. la validation des données saisies.

1. Parcours logique de saisie des données

Comparons par exemple la logique de saisie des données en cas: 1. de données groupées par valeur; 2. de données brutes relatives à deux échantillons. Pour chacun de ces deux cas, le nombre de colonnes de données à saisir est égal à 2:

Cas 1: col. 1 = valeurs d'une variable

col. 2 = fréquences associées à ces valeurs

Cas 2: col. 1 = observations relatives au premier échantillon

col. 2 = observations relatives au deuxième échantillon

Les masques de saisie correspondant à chacun de ces cas sont représentés en figures 8.3 et 8.4.

i	Valeurs x_i	Fréquences f_i
1		
2		
3		
4		
5		
6		
7		

Figure 8.3: Masque de saisie pour des données groupées par valeur

i	Echantillon 1	Echantillon 2
1		
2		
3		
4		
5		
6		
7		

Figure 8.4: Masque de saisie pour des données brutes relatives à deux échantillons

Bien que le masque de saisie soit identique (si l'on excepte l'en-tête de chaque colonne), la logique du parcours de saisie est différent. Pour le premier cas, il semble naturel d'entrer successivement chaque valeur suivie de la fréquence qui lui est associée. Comme le montre la figure 8.5, on parcourt alors successivement les deux colonnes de chaque ligne.

Le parcours logique de saisie de données dans le deuxième cas n'est pas identique. Il ne serait pas naturel d'entrer la première observation de l'échantillon 1, la première observation de l'échantillon 2, puis la deuxième observation de l'échantillon 1, la deuxième observation de l'échantillon 2, et ainsi de suite.

Dans ce cas, comme le montre la figure 8.6, il est logique de saisir d'abord toutes les observations de l'échantillon 1, puis de passer à la colonne suivante afin de saisir les observations de l'échantillon 2.

i	Valeurs x_i	Fréquences f_i
1		
2		
3		
4		
5		
6		
7		

Figure 8.5: Parcours logique de saisie de données groupées par valeur

i	Echantillon 1	Echantillon 2
1		
2		
3		
4		
5		
6		
7		

Figure 8.6: Parcours logique de saisie de données brutes relatives à deux échantillons

Cet exemple, présenté pour deux types particuliers de données, montre bien que si l'on veut travailler dans un environnement assisté, il n'est pas possible de prévoir une procédure unique de saisie des données. Chaque type de données doit être soutenu par un processus de saisie adéquat.

2. Validation des données saisies

Le type de traitement étant connu lors de la procédure de saisie des données, celle-ci peut être responsable d'une première vérification de la validité des entrées. Certaines vérifications sont dites générales, car elles sont applicables lors de la saisie de n'importe quel type de données. Elles concernent essentiellement l'étendue des valeurs permises, en fonction des contraintes techniques de la machine.

D'autres vérifications dépendent directement du type de données saisies, ou du type de traitement auquel ces données seront soumises. Celles-ci concernent par exemple:

- l'intervalle des valeurs permises par le type spécifique de traitement qui sera appliqué aux données (par exemple, une probabilité doit être comprise entre 0 et 1);
- le signe des données (par exemple, une fréquence ne doit pas être négative);
- la cohérence des bornes lors d'une saisie de données groupées en classes.

Les deux problèmes spécifiques que nous venons de soulever montrent bien qu'il est nécessaire de répertorier tous les types de données possibles afin de pouvoir les traiter de manière distincte.

Le nombre de types dépend bien évidemment de l'ensemble des outils développés. Certains outils peuvent offrir la possibilité de traiter plusieurs types, d'autres uniquement un type particulier.

L'étude des particularités propres aux données relatives à chaque outil nous a amené à définir les 13 types suivants:

1. données brutes (ou individuelles)
2. données groupées par valeur
3. données groupées par classe
4. données brutes avec poids associés
5. données groupées par valeur avec poids associés
6. données groupées par classe avec poids associés
7. données catégoriques avec label associé
8. données relatives à un échantillon (ou à une série temporelle)
9. données relatives à deux échantillons (ou à deux variables)
10. données relatives à trois échantillons
11. données relatives à quatre échantillons
12. données sous forme de table de contingence
13. données pour la classification

Le type de données détermine totalement l'environnement de saisie:

- le parcours de saisie;
- la vérification de la validité des données saisies;
- les caractéristiques de chaque colonne, une colonne pouvant représenter par exemple:
 - des labels associés à chaque observation,
 - les valeurs d'une variable quantitative,
 - des fréquences associées à ces valeurs,
 - des poids associés à ces valeurs,
 - les bornes inférieures ou supérieures des classes.
- le nombre de variables (et éventuellement d'observations) à saisir;
- les en-têtes relatifs à chaque colonne.

La saisie des données proprement dite n'est que la première tâche de la procédure Proc_Donnees. En fonction de règles définies pour chaque type, celle-ci permet également d'effectuer les opérations suivantes:

- la modification (correction et ajout) de données, qui est strictement soumise aux mêmes règles que la saisie des données;
- la suppression de données. En fonction du type de données, il peut être pertinent de supprimer une ligne entière, une colonne entière, ou un élément particulier du tableau de données;
- la sauvegarde des données dans un fichier;
- la récupération d'un fichier préalablement sauvegardé. La récupération des données n'est autorisée que lorsque celles-ci ont été sauvegardées par un outil compatible au niveau de son type.

Comme nous l'avons déjà mentionné, les outils statistiques n'ont pas la vocation de répondre à des besoins professionnels d'utilisation des méthodes statistiques, mais au contraire à celle d'aide à l'apprentissage et à la compréhension des notions.

Il n'est pas utile dans un tel contexte d'offrir des outils permettant de traiter des ensembles de données de taille illimitée. Il est préférable de limiter la capacité (par exemple 100 observations, ou 100 lignes) et de porter davantage d'efforts sur la simplicité et la clarté de l'interface de saisie des données.

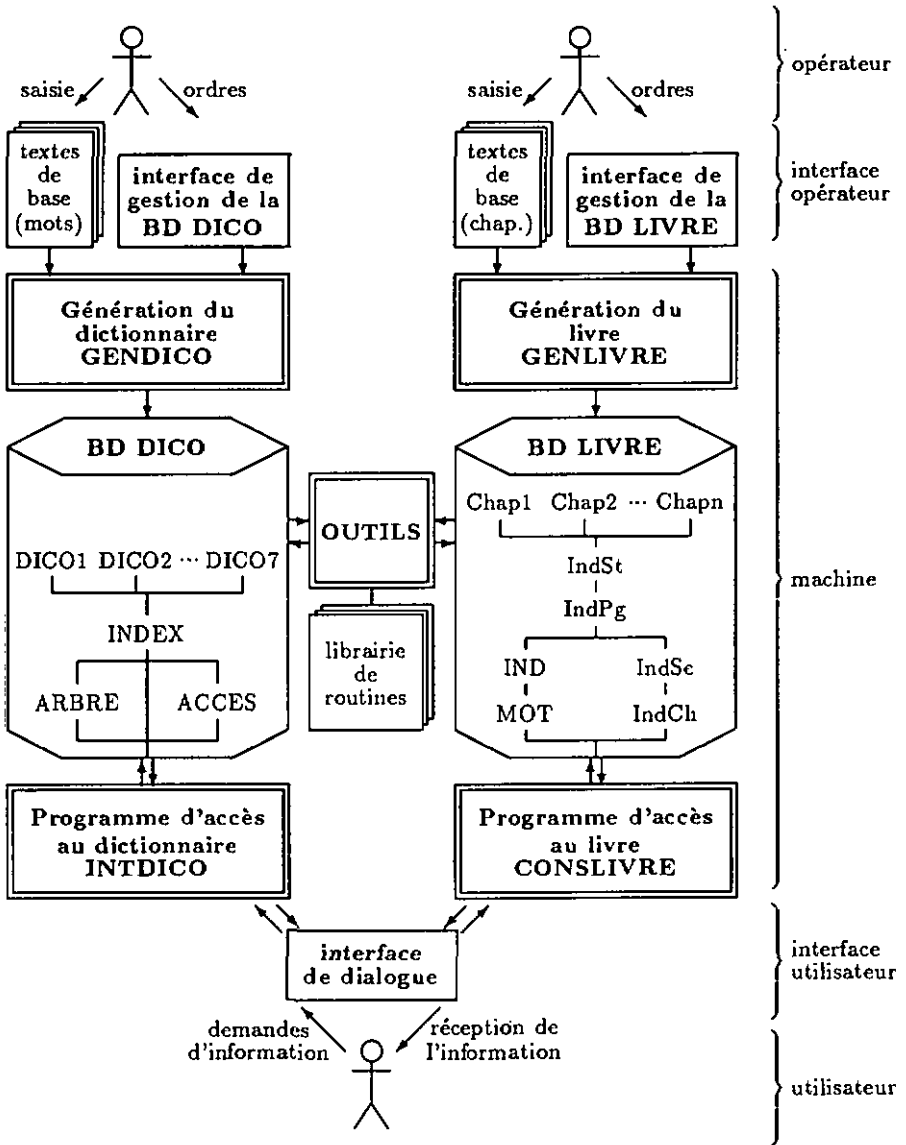
Conclusions

Par les principes que nous avons exposés, nous avons posé les bases nécessaires au développement d'un outil de Formation Assistée par Ordinateur permettant une navigation aisée entre un livre, un dictionnaire et des méthodes. Certes, la juxtaposition entre du texte et des méthodes n'est pas nouvelle. Toutefois, l'intégration d'un dictionnaire de taille réelle présente une innovation certaine. Une autre originalité du concept exposé est l'intégration en un seul outil de trois niveaux d'apprentissage: un apprentissage structuré par la consultation d'un livre; un apprentissage ponctuel par l'accès à un dictionnaire; un apprentissage pratique par la mise à disposition de méthodes, ou d'outils.

L'objectif de cette thèse a été de dépasser la formulation de concepts, afin de proposer une méthode pratique de réalisation d'un outil de formation, tenant compte de l'importance des aspects pédagogiques liés au problème, et proposant une méthodologie de développement informatique.

Le système informatique complet peut se synthétiser selon le schéma proposé ci-dessous. Sur la base de ce schéma, nous pouvons mettre en évidence les principales caractéristiques de la méthode proposée:

- La création d'un système de formation assistée par ordinateur consiste à établir un pont entre deux mondes différents: d'une part le monde de l'étudiant (l'utilisateur de l'outil), qu'il est nécessaire de cerner, de comprendre, afin de fournir une interface qui corresponde à ses attentes et à son niveau de compréhension; et d'autre part le monde du professeur, de l'expert, qui par sa connaissance et sa maîtrise du domaine, sera capable de déterminer et de gérer le contenu, le savoir à transmettre.
- Un dictionnaire et un livre ne sont pas structurés de la même manière. Un système intégrant les deux types de modules aura donc deux bases de données distinctes, structurées en fonction de leurs spécificités propres.
- Le système comprend deux interfaces opérateur, l'opérateur étant la personne chargée de la gestion du contenu de chaque base de données.



- Le système ne comprend qu'une seule interface utilisateur, qui ouvre l'accès aux deux modules principaux qui sont la lecture du livre et la consultation du dictionnaire.
- Les outils forment un module séparé, accessible indifféremment par le dictionnaire ou par le livre. Il serait également possible d'envisager une entrée directe vers les outils, supprimant ainsi le passage obligé par un des deux autres modules.

La littérature dans le domaine de l'Enseignement Assisté par Ordinateur est souvent basée sur des exposés théoriques de certains aspects spécifiques du problème. Nous avons voulu ici concevoir et implanter un système de taille réelle, afin de fournir un ensemble cohérent, fondé sur une expérience vécue.

Cette expérience nous a démontré la faisabilité d'un tel système. Les concepts que nous proposons ont été testés sur une application dans le domaine de l'apprentissage de la statistique. Ces concepts ne sont toutefois pas indissociablement liés à ce domaine d'apprentissage particulier, et pourraient être étendus à une multitude d'autres domaines.

Bibliographie

- [1] Balpe J.P. (1990). *Les hyperdocuments - hypertextes, hypermédias*, Eyrolles, Paris.
- [2] Barbey G. (1971). *L'enseignement assisté par ordinateur*, Casterman, Tournai.
- [3] Barlow M. (1987). *Formuler et évaluer ses objectifs en formation*, Chronique sociale, Lyon.
- [4] Barthet M.F. (1985). *Conception du logiciel d'interface homme-machine*, Ergonomie et conditions de travail en informatique, édité par René Patesson, Editions de l'Université de Bruxelles.
- [5] Besnainou R., Muller C. et Thouin Ch. (1988). *Concevoir et utiliser un didacticiel*. Les Editions d'Organisation, Paris.
- [6] Birzea C. (1982). *La pédagogie du succès*, Presses Universitaires de France, Paris.
- [7] Conklin J. (1987). *Hypertext: an Introduction and Survey*, IEEE Computer, 20(9), 17-41.
- [8] Crossley K. et Green L. (1990). *Le design des didacticiels*, Editions Art-Culture-Lecture, Paris.
- [9] Crowder N.A. (1963). *Intrinsic Programming: Facts, Fallacies and Future*, Prospectives in Programming, édité par R.T. Filep, The Macmillan Company, New York.
- [10] Crowder N.A. (1966). *L'arithmétique des ordinateurs*, (*The Arithmetic of Computers*), version française de Jacques Matalon, Cours d'instruction programmée TutorText, Editions Gamma, Paris.

- [11] Davies I. (1971). *The management of learning*, Mc Graw-Hill, Londres.
- [12] Dodge Y. (1993). *Encyclopédie et Dictionnaire en statistique*. Editions Dunod, Paris.
- [13] Dodge Y. et Rebetez N. (1992). *Computer-Aided Dictionary of Statistical Terms*, SoftStat'91, édité par F.Faulbaum, Gustav Fischer, 515-522.
- [14] Duchateau Ch. (1986). *Initier à l'informatique: réduire à la maîtrise d'une technique ou promouvoir un humanisme*, Technologies de l'information: aspects humains et sociaux, édité par Luc Wilkin, Editions de l'Université de Bruxelles.
- [15] Fitz-Gibbon C. (1988). *How to design a program evaluation*, Sage, Londres.
- [16] Gagné R. (1976). *Les principes fondamentaux de l'apprentissage: application à l'enseignement*, Editions HRW, Montréal.
- [17] Goodwin D.L. et Coates T.J. (1976). *Helping students help themselves*, Prentice-Hall, Englewood Cliffs.
- [18] Hartley J. (1978). *Designing instructional text*, Kogan Page, Londres.
- [19] Hennart M., Berthon J.F. et Binse M. (1986). *Ecrire des logiciels pédagogiques*, Cedic/Nathan, Paris.
- [20] Holland J.G. (1976). *Analysis of behavior in planning instruction*, Addison-Wesley, Reading.
- [21] Houziaux M.O. (1972). *Vers l'enseignement assisté par ordinateur*. Presses Universitaires de France, Paris.
- [22] Jones H.R. (1962). *Color coding*, Human Factors, 4, 6, 355-366.
- [23] Keer S.T. (1990). *Wayfinding in an Electronic Database: the Relative Importance of Navigational Cues vs. Mental Models*, Information Processing & Management, 26(3), 437-447.
- [24] Klare G. et al. (1955). *The relationship of patterning (Underlying) to immediate retention and to acceptability of technical material*, Journal of Applied Psychology, 39, 1, 40-42.
- [25] Kotz S. et Johnson N.L. (1985). *Encyclopedia of Statistical Sciences*, John Wiley & Sons.

- [26] Kowal J.L. (1986). *La technique du fenêtrage comme moyen facilitant le dialogue homme-machine dans une expérience d'une leçon de formation assistée par ordinateur*, L'homme et l'écran, aspects de l'ergonomie en informatique, édité par René Patesson, Editions de l'Université de Bruxelles.
- [27] Kruskal W.H. et Tanur J.M. (1978). *International Encyclopedia of Statistics*, The Free Press, Macmillan Publishing Co., New York.
- [28] Leplat J. (1972). *La psychologie du travail en ergonomie*, Traité de psychologie appliquée, Vol. 3, Presses Universitaires de France.
- [29] Mager R. (1988). *Comment définir des objectifs pédagogiques*, Gauthier-Villars, Paris.
- [30] Marchionini G. (1989). *Making the Transition from Print to Electronic Encyclopaedias: Adaptation of Mental Models*, International Journal of Man-Machine Studies, 30(6), 591-618.
- [31] Mariott F.H.C. (1990). *A Dictionary of Statistical Terms*, publié pour l'International Statistical Institute par Longman Scientific & Technical et John Wiley & Sons.
- [32] Martin J.P. (1987). *Du bricolage à l'industrialisation: La qualité des logiciels*, AFNOR, Paris.
- [33] McCracken D.L. et Akscyn R. (1984). *Experience with the ZOG Human-Computer Interface System*, International Journal of Man-Machine Studies, 21(4), 293-310.
- [34] Miller G.A. (1968). *Psychology and information*, American Documentation, 193, 286-289.
- [35] Miller L.A. et Thomas G.C. (1977). *Behavioral issues in the use of interactive systems*, International Journal of Man-Machine Studies, Vol. 9, No 5.
- [36] Montmollin M. de (1970). *L'enseignement programmé*. Collection "que sais-je?", Presses Universitaires de France, Paris.
- [37] Montmollin M. de (1986). *L'ergonomie*. Editions La Découverte, Paris.
- [38] Ohlsson S. (1985). *Some Principles of Intelligent Tutoring*, Learning Research and Development Center, University of Pittsburgh.

- [39] O'Neil H.F. et Dansereau D. (1988). *Learning strategies*, Academic Press, Londres.
- [40] Popham W.J. et Baker E.L. (1981). *Comment programmer une séquence pédagogique*, Bordas, Paris.
- [41] Roger D., Binot Ch., Beyaert G. et Willaëys. (1986). *Dialogue homme-machine et communication graphique: le problème de l'évaluation*, L'homme et l'écran, aspects de l'ergonomie en informatique, édité par René Patesson, Editions de l'Université de Bruxelles.
- [42] Romiszowski A.J. (1986). *Developing auto-instructional materials: from programmed text to CAL and interactive video*, Kogan Page, Londres.
- [43] Ronchetti E., Antille G. et Polla M. (1989). *STEP I, STatistique et Probabilités: une Introduction*, Presses Académiques, Neuchâtel.
- [44] Sackman H. (1970). *Man-Computer Problem Solving*, Auerbach Publishers, Princeton.
- [45] Savoy J. (1988). *Le livre électronique Ebook3*, Technique et Science Informatiques, 7(5), 437-450.
- [46] Scholer M. (1983). *La technologie de l'éducation: concept, bases et applications*, Presses de l'Université, Montréal.
- [47] Shneiderman B. (1987). *Designing the User Interface, Strategies for Effective Human-Computer Interaction*, Addison-Wesley Publishing Company.
- [48] Skinner B.F. (1968). *La révolution scientifique de l'enseignement*, (*The Technology of Teaching*), traduit de l'américain par A. Richelle. Charles Dessart, Editeur, Bruxelles.
- [49] Smith S.L. (1963). *Color coding and visual separability in information displays*, Journal of Applied Psychology, 47, 358-364.
- [50] Sommerville I. (1985). *Software Engineering*, second edition. Addison-Wesley Publishing Company, Inc.
- [51] Sperandio J.C. (1985). *L'ergonomie de l'informatique, évolution passée et perspective de recherche*, Ergonomie et conditions de travail en informatique, édité par René Patesson, Editions de l'Université de Bruxelles.
- [52] Talyzina N.F. (1980). *De l'enseignement programmé à la programmation de la connaissance*, Presses universitaires, Lille.

- [53] Tracey W. (1971). *Designing training and development systems*, American Management Association, New York.
- [54] Weinberg G.M. (1971). *The Psychology of Computer Programming*, Van Nostrand Reinhold Company Inc.
- [55] Yankelovich N., Meyrowitz N. et van Dam A. (1985). *Reading and Writing the Electronic Book*, IEEE Computer, 18(10), 15-30.
- [56] Yankelovich N., Haan B., Meyrowitz N. et Drucker S. (1988). *Intermedia: the Concept and Construction of a Seamless Information Environment*, IEEE Computer, 21(1), 81-96.
- [57] Weyer S.A. et Borning A.H. (1985). *A Prototype Electronic Encyclopedia*, ACM Transactions on Office Information Systems, 3(1), 63-88.
- [58] *Introduction to Computer-Based Education*, deuxième édition, (1985), Digital Equipment Corporation.
- [59] *Prise en main de Framework III*, (1988), La Commande Electronique.
- [60] *Turbo Pascal: manuel de l'utilisateur*, sixième édition, (1988), Borland International.
- [61] *Turbo Pascal: manuel de référence*, sixième édition, (1988), Borland International.