

Sc.N.74/53^B

SYSTEME INFORMATIQUE DE DOCUMENTATION
POUR L'ASSISTANCE CARDIO-CIRCULATOIRE.

THESE

pour l'obtention du

DOCTORAT DE SPECIALITE
MATHÉMATIQUES APPLIQUEES (3^o Cycle)

Soutenu le 22 Juin 1974

par

YVES SCHNEPF

JURY : - M. C. PAIR
- Mme M. CREHANGE
- M. R. BENICHOUX
- M. J.C. DERNIAME

Président
Examineurs



UNIVERSITE DE NANCY I

U. E. R. DE SCIENCES MATHÉMATIQUES

SYSTEME INFORMATIQUE DE DOCUMENTATION
POUR L'ASSISTANCE CARDIO-CIRCULATOIRE.

THESE

pour l'obtention du

DOCTORAT DE SPECIALITE
MATHÉMATIQUES APPLIQUÉES (3^o Cycle)

Soutenu le 22 Juin 1974

par

YVES SCHNEPF

JURY : _ M. C. PAIR
_ Mme M. CREHANGE
_ M. R. BENICHOUX
_ M. J.C. DERNIAME

Président
Examineurs



Je remercie en tout premier lieu, et très vivement, Monsieur PAIR, Directeur de l'Institut Universitaire de Calcul Automatique, d'avoir bien voulu me faire l'honneur de présider ce jury.

Mes remerciements s'adressent aussi à Monsieur BENICHOUX, Directeur de l'Institut de Recherches Chirurgicales, qui a été à l'origine de ce travail.

C'est grâce à Madame CREHANGE, à ses encouragements, à ses conseils, et à l'attention constante qu'elle a su me prodiguer que cette recherche a pu être réalisée ; je lui en témoigne ici ma profonde gratitude.

Je remercie également Monsieur DERNIAME pour l'honneur qu'il me fait de participer à ce jury.

A Monsieur SZTUKA, j'exprime toute ma reconnaissance pour l'aide sympathique qu'il m'a apportée au cours de cette étude.

Enfin, je remercie toutes les personnes qui ont participé à la réalisation matérielle de cette thèse.

S O M M A I R E

	Pages
<u>CHAPITRE I - INTRODUCTION</u>	
<u>I. 1 - PRESENTATION DU PROBLEME</u>	1.1
<u>I. 2 - CHAINE DOCUMENTAIRE</u>	1.3
<u>CHAPITRE II - DESCRIPTION DES DOCUMENTS</u>	
<u>II. 1 - INTRODUCTION</u>	2.1
<u>II. 2 - LANGAGE DE DESCRIPTION DES DOCUMENTS</u>	2.3
A - Notion de carte syntaxique	2.3
B - Cartes syntaxiques du langage de description	2.5
B - 1. Document	2.5
B - 2. Commentaire	2.5
B - 3. Identification	2.7
B - 4. Liste de paragraphes	2.9
B - 5. Liste de vecteurs	2.9
B - 6. Liste de composantes	2.11
B - 7. Liste de valeurs	2.11
B - 8. Groupe adjectif	2.15
<u>II. 3 - UTILISATION DU LANGAGE DE DESCRIPTION</u>	2.19
A - Description des expériences	2.19
B - Description des tableaux	2.24
C - Description de l'échelle des temps	2.27
D - Utilisation des adjectifs	2.31
E - La composante MODIF	2.36
<u>II. 4 - PRESENTATION D'UN BORDEREAU DE PERFORATION</u>	2.40
A - Numérotage	2.40
B - Partie identification	2.40
C - Partie Liste de paragraphes	2.40
D - Partie Commentaire	2.42

II. 5 - EXEMPLE DE DESCRIPTION D'UN DOCUMENT

Pages
2.43

CHAPITRE III - CHAINE D'ACQUISITION DES DOCUMENTS

III. 1 - INTRODUCTION

3.1

III. 2 - DIRECTIVES D'UTILISATION DE LA CHAINE

3.3

A - Les tâches à entreprendre par l'utilisateur

3.3

B - Description des tâches

3.5

B - 1. Décrire les documents et les perforer

3.5

B - 2. Demander l'activation du programme de traduction

3.5

B - 3. Dépouiller le listing produit par le programme TRD

3.5

III. 3 - LES FICHIERS.

3.10

A - Le lexique des noms de paragraphes et de composantes (FLPC)

3.10

A - 1. Généralités

3.10

A - 2. Description du fichier FLPC

3.11

A - 3. Création du fichier FLPC

3.12

B - Le lexique des mots

3.14

B - 1. Généralités

3.14

B - 2. Méthode de classement par ordre alphabétique à l'aide d'un arbre binaire

3.15

B - 3. Relation de hiérarchie

3.20

B - 4. Description du fichier FLM

3.23

B - 5. Description du fichier FH

3.25

B - 6. Exemple de rangement de mots et de leur structure hiérarchique

3.26

B - 7. Création et mise à jour du fichier FLM

3.29

B - 8. Création et mise à jour du fichier FU

3.31

C - Les fichiers des documents (FDOC et FDOCD)

3.36

C - 1. Généralités

3.36

C - 2. Forme interne des listes de paragraphes

3.36

C - 3. Description des fichiers FDOC et FDOCD

3.44

	Pages
<u>III. 4 - REALISATION</u>	3.48
<u>CHAPITRE IV - INTERROGATION DU FICHER DES DOCUMENTS.</u>	
<u>IV. 1 - INTRODUCTION</u>	4.1
<u>IV. 2 - TABLE DE SELECTION ET DE TRAITEMENT</u>	4.3
A - Subdivision "texte de la question"	4.3
A - 1. Description et règles d'écriture	4.3
A - 2. Vecteur pertinent	4.6
A - 3. Valeurs "en accord"	4.8
B - Subdivision "ordre de sélection"	4.10
B - 1. Case marquée	4.11
B - 2. Sélection de vecteurs d'une même expérience	4.12
B - 3. Colonne de sélection pertinente	4.12
B - 4. Réponses à une colonne de sélection	4.14
C - Subdivision "Traitements" et "ordres de traitements"	4.14
D - Fonctionnement de la table de sélection et de traitements	4.15
E - Les traitements	4.18
E - 1. Instructions de gestion de la table de sélection	4.18
E - 2. Instruction de répétition d'une liste d'instructions	4.18
E - 3. Autres instructions	4.19
<u>IV. 3 - ENTREE D'UN TRAIN DE QUESTIONS EN MACHINE</u>	4.21
A - Identification de la question	4.21
B - Limites d'exploitation du fichier des documents	4.21
C - Sélection sur l'entête	4.22
C - 1. Sélection sur la référence	4.22
C - 2. Sélection sur l'année de parution	4.23
D - Présentation de la table de sélection et de traitement sur les bordereaux	4.24
E - Train de questions	4.25

	Pages
<u>IV. 4 - REALISATION DU SYSTEME D'INTERROGATION</u>	4.26
A - Création d'un fichier de questions codées (FQST)	4.26
B - Exploitation du fichier des documents à partir des questions codées	4.27
C - Exemple	4.27

CONCLUSION

BIBLIOGRAPHIE

CHAPITRE I

INTRODUCTION

I. 1 - PRESENTATION DU PROBLEME.

Le but de ce travail est de réaliser un système de documentation concernant l'Assistance Cardio-Circulatoire sur la demande^{de} l'Institut de Recherches Chirurgicales (CHU de Brabois - Vandoeuvre-les-Nancy. Directeur : Pr. R. BENICHOUX).

Nous introduisons notre propos par quelques mots sur l'Assistance Cardio-Circulatoire.

Le but de cette technique est de prendre en charge une part plus ou moins grande du travail d'un coeur défaillant pendant le temps nécessaire au rétablissement d'une fonction myocardique satisfaisante. Il s'agit de mettre en oeuvre des procédés mécaniques susceptibles d'augmenter le débit ou/et la pression systémique sans pour autant accroître les besoins énergétiques du myocarde.

De nombreux systèmes ont été réalisés et expérimentés sur l'animal (chiens, porcs, veaux) et sur l'homme avec des succès variables. Les systèmes diffèrent suivant qu'ils fonctionnent en régime continu ou pulsé, avec ou sans dérivation du sang, avec ou sans synchronisation. Des appareillages variés sont mis en oeuvre tels que des pompes, des oxygénateurs, des échangeurs thermiques, des ballonnets intra-aortiques, etc... Le bon fonctionnement de l'assistance dépend en grande partie du réglage de ces appareils.

L'assistance cardio-circulatoire est employée dans diverses circonstances telles que des insuffisances ventriculaires (gauches ou droites), des chocs hypovolémiques ou cardiogéniques, des syndromes de bas débit cardiaque après circulation extra-corporelle, etc...

Il est difficile d'appréhender dans son ensemble les problèmes de l'assistance cardio-circulatoire étant donné le grand nombre d'articles publiés tant en France qu'à l'étranger, la diversité des procédés expérimentés, la multiplicité des conditions d'application.

Les personnes intéressées par les problèmes de l'assistance ont ressenti la nécessité d'avoir à leur disposition un outil de travail qui puisse les aider à répondre à une question précise à bref délai et sans avoir à consulter manuellement la littérature existante.

Exemples de questions.

- Dans des conditions pathologiques et avec une méthode d'assistance déterminées, quel est le taux de mortalité obtenu sur l'animal ? sur l'homme ?

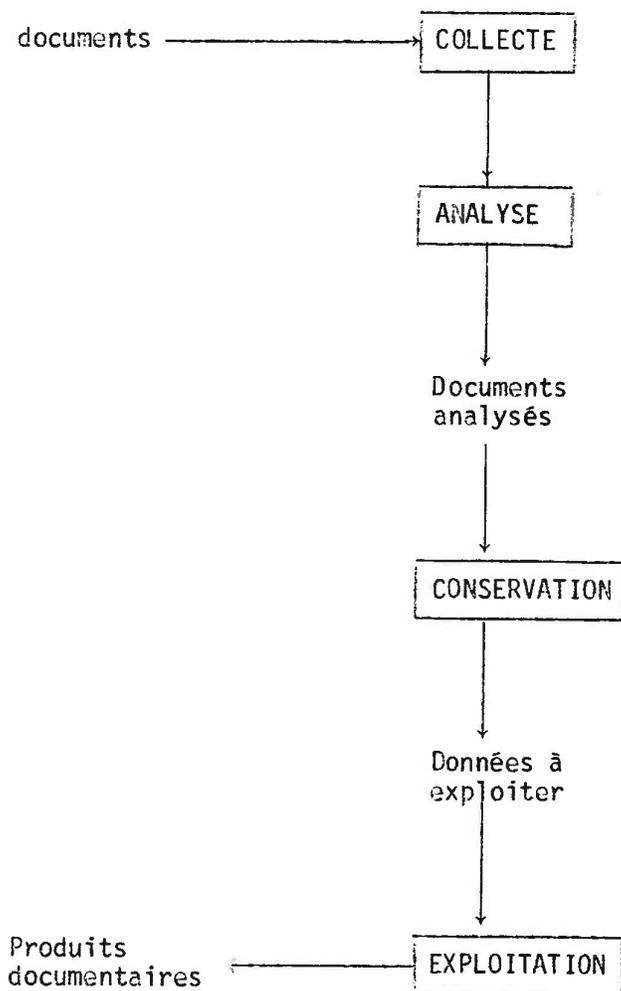
- Dans ces mêmes conditions, quelle est la valeur maximale observée d'un certain paramètre ?

Les réponses à ces questions aident les chercheurs à déterminer les meilleures méthodes à employer dans des conditions déterminées, à définir les paramètres les plus importants, à diriger les nouvelles recherches dans les voies intéressantes, etc...

Le système de documentation que nous mettons au point doit donc permettre non seulement de sélectionner des documents qui répondent à certains critères mais aussi de faire des traitements sur leur contenu (calculs de moyenne, de maximum, de minimum, de pourcentage, etc...).

II. 2 - CHAÎNE DOCUMENTAIRE.

Nous voulons aboutir à une chaîne documentaire dont le schéma très général est donné ci-dessous.



Un système documentaire apporte une solution aux différentes tâches encadrées sur ce schéma, certaines pouvant être traitées par des moyens informatiques. Nous allons étudier très rapidement ces diverses étapes en indiquant les solutions que nous avons retenues pour notre propre système.

A - Collecte des documents.

Cette phase de la chaîne documentaire est manuelle, longue, coûteuse et souvent malaisée. Elle demande une organisation solide si l'on veut tendre vers une collecte quasi-exhaustive des documents qui concernent le domaine d'application retenu par le système documentaire.

En ce qui nous concerne, ce maillon de la chaîne n'a pas été organisé, mais devra l'être dans l'hypothèse d'un élargissement du cadre des utilisateurs (d'autres CHU par exemple).

B - Analyse du contenu.

Cette phase est la plus importante de la chaîne et est largement conditionnée par les vœux des utilisateurs quant à l'exploitation ultérieure que l'on veut faire des documents et par la décision d'utiliser des moyens informatiques ou non pour le stockage des informations. Le travail d'analyse est souvent difficile et conditionne la qualité de l'exploitation que l'on fera des documents.

Une méthode d'analyse et de description des documents courante est l'utilisation de mots-clé (ou descripteurs). Le système MEDLARS (Medical Literature Analysis and Retrieval System) mis au point par la National Library of Medicine de Bethesda (U.S.A.) en offre un exemple connu dans les milieux médicaux. Les descripteurs sont souvent répertoriés dans un thesaurus de manière conceptuelle, c'est-à-dire par affinités sémantiques et sont reliés par des notions de hiérarchie et d'association.

Nous avons quant à nous retenu une méthode d'analyse un peu plus fine afin de pouvoir répondre aux vœux des utilisateurs (voir I. 1). La méthode de description que nous avons mise au point est inspirée au départ par le travail que Mme de Lary a fait pour traiter des fiches de relevés géologiques à la demande du CRPG (Centre de Recherches Pétrographiques et Géochimiques - Nancy) et qui fait l'objet d'une thèse de spécialité.

Le stockage des documents analysés par des moyens informatiques exige que leur description soit très concise autant dans le fond que dans la forme. D'où la définition d'un langage de description des documents auquel nous consacrons tout le chapitre II.

C - Conservation des documents analysés.

Les documents analysés et décrits sont stockés automatiquement en machine à partir de cartes perforées en ce qui nous concerne. Cette opération est décrite dans le chapitre III. Nous y décrivons aussi les lexiques qui sont nécessaires pour effectuer ce travail.

D - Exploitation.

Deux modes d'exploitation sont généralement mis en oeuvre dans les systèmes documentaires.

Le premier est la création d'index envoyés périodiquement aux utilisateurs (diffusion générale). Une diffusion plus fine consiste à éditer des index personnalisés en accord avec le profil scientifique de l'utilisateur (diffusion sélective).

Le deuxième mode d'exploitation est la recherche rétrospective, c'est-à-dire la recherche de documents parmi ceux qui existent et qui répondent à des critères généralement exprimés sous forme d'expressions logiques de descripteurs.

C'est ce deuxième mode d'exploitation que nous avons adopté pour notre système dont la vocation première est la recherche rétrospective sur des critères de sélection (plus affinés que ceux permis par les descripteurs). En outre, nous voulons pouvoir faire des traitements sur les informations sélectionnées. Nous appelons interrogation cette exploitation des documents (chapitre IV).

E. - Réalisation.

La conception de ce système a demandé un long travail de concertation entre médecins et informaticiens, les premiers n'ayant pas de cahier des charges à soumettre aux seconds. D'autre part on conçoit les difficultés soulevées par des expressions telles que "mode d'évolution de" ou "influence de... sur..." usitées dans les textes médicaux.

La programmation a été faite en langage COBOL - CII et les programmes ont été testés sur la machine CII-10 070 de l'IUCA (Institut Universitaire de Calcul Automatique - Château du Montot - 54500 Vandoeuvre-les-Nancy. Directeur : Monsieur PAIR).

CHAPITRE II

DESCRIPTION DES DOCUMENTS

II. 1 - INTRODUCTION.

Les documents sont décrits à l'aide d'un langage de description dont la mise au point n'a pu être faite que grâce à la collaboration de M. SZTUKA compétent en matière médicale. Ce travail a été long et a demandé que s'installe un dialogue presque constant entre médecin et informaticien. Nous avons essayé de définir un langage dont la syntaxe soit facilement assimilable par un médecin. Nous allons dégager les notions fondamentales de ce langage.

L'information : "La quantité de sodium est de 140" sera transcrite de la manière suivante dans notre langage :

NA = 140

Remarquons tout de suite que les unités sont implicites. (mEq/l pour NA).

Une telle information est souvent une composante d'une information plus complexe. D'où la dénomination de composante pour une telle information (NA est le nom de la composante). Prenons l'exemple : "Trois chiens de cinq ans d'âge". Nous écrirons :

ESPECE = CHIEN ;
AGE = 5 ;
NOMBRE = 3 #

Cet ensemble de composantes est appelé vecteur. Notons que la signification d'un vecteur ne dépend pas de l'ordre d'écriture de ses composantes. On aurait tout aussi bien pu commencer par donner le nombre ou l'âge des chiens.

Les différents vecteurs qui décrivent un document sont ventilés dans des rubriques appelées des paragraphes. Le vecteur précédent est rangé dans le paragraphe dont le nom est PATIENT.

PATIENT = ESPECE = CHIEN ;
AGE = 5 ;
NOMBRE = 3 /

Un paragraphe peut regrouper un ou plusieurs vecteurs. Si en plus des chiens il y a aussi un porc, le paragraphe devient :

PATIENT = ESPECE = CHIEN ;
AGE = 5 ;
NOMBRE = 3
;#
ESPECE = PORC
NOMBRE = 1 /

Il est à remarquer que les deux vecteurs qui composent ce paragraphe n'ont pas le même nombre de composantes.

Un document va être décrit par une suite de paragraphes. Il se pose un problème de liaison entre les vecteurs des différents paragraphes. En d'autres termes, il faut pouvoir exprimer que tel vecteur du paragraphe P1 "va avec" tel autre vecteur du paragraphe P2. Si pour les patients décrits précédemment le document donne plusieurs séries de résultats, chaque série sera décrite par un vecteur. Ces vecteurs sont rangés dans le paragraphe RESULTAT. Il faut alors pouvoir faire le lien entre les résultats qui concernent les chiens et ceux qui concernent le porc. Ce point sera développé en II. 3 - A.

Nous devons également pouvoir donner des descriptions de tableaux qui figurent dans les documents (Par exemple, des tableaux de résultats).

Avant de répondre à ce problème, nous avons voulu fixer les règles de syntaxe de notre langage (II. 2). Puis nous expliquons comment utiliser ce langage (II. 3), c'est-à-dire que nous l'abordons sous l'angle sémantique.

Ce chapitre doit être lu attentivement par l'utilisateur qui veut décrire des documents et constitue en quelque sorte une notice explicative. En outre, nous donnons en II. 4 le format d'un bordereau de perforation (les documents sont perforés sur cartes pour leur entrée en machine) et en II. 5, le lecteur trouvera un exemple de description de document.

II. 2 - LANGAGE DE DESCRIPTION DES DOCUMENTS.

A - Notion de carte syntaxique.

Nous avons choisi d'utiliser les cartes syntaxiques pour donner les règles de syntaxe de notre langage. Celles-ci nous semblent être d'une lecture plus facile pour les utilisateurs que les notations de Bacchus. Cette notation a d'ailleurs été utilisée dans les rapports d'Algol 60 et ne présente pas un caractère original.

Une carte syntaxique est destinée à donner les règles d'écriture d'un langage à l'aide d'un dessin constitué :

- de cercles qui contiennent les symboles de base du langage,
- d'ovales qui contiennent des notions définies à cet endroit,
- de rectangles qui contiennent des notions définies ailleurs,
- de flèches horizontales qui signifient : "suivi de",
- de flèches descendantes qui signifient : "peut être".

Exemple de carte syntaxique (voir schéma n° 1, page 2.4)

Cette carte exprime que la notion V peut être :

- soit le symbole de base U,
- soit la notion W suivie de U.

Mais V ne peut pas être la notion W toute seule. Remarquons que la notion W est définie ailleurs puisqu'elle se trouve dans un rectangle.

Carte syntaxique du nombre entier (voir schéma n° 2, page 2.4)

La carte syntaxique (schéma n° 2) doit être lue de la manière suivante :

Un nombre entier est :

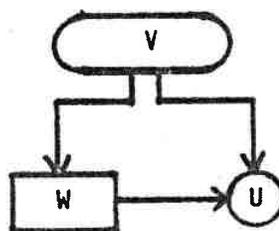
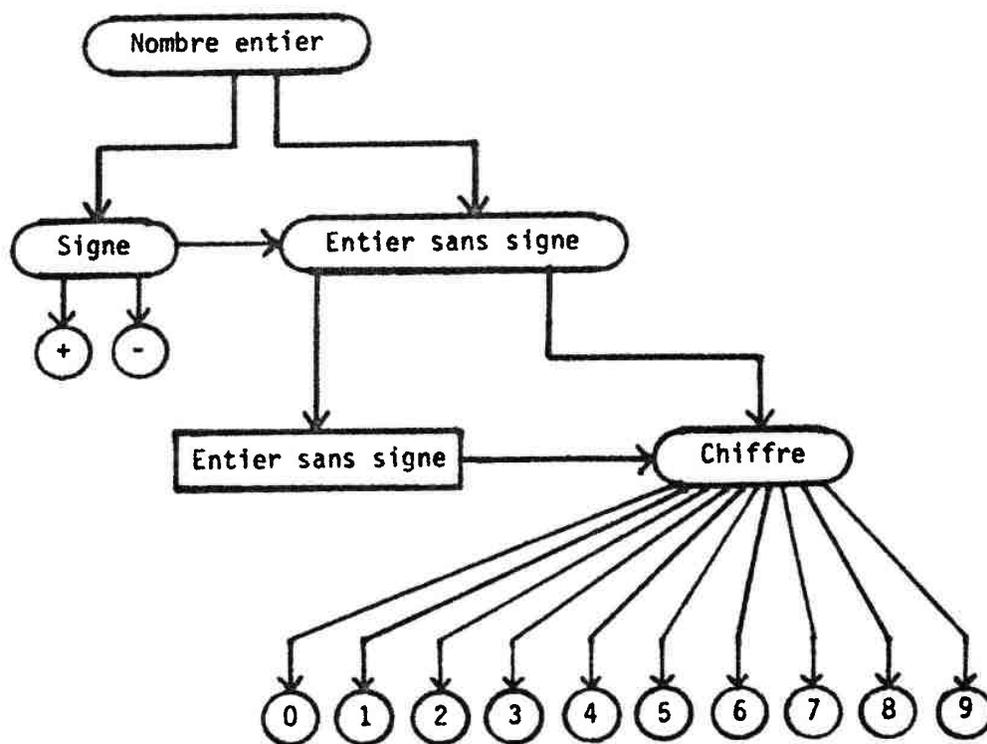
- soit un entier sans signe,
- soit un signe suivi d'un entier sans signe.

Un entier sans signe est :

- soit un chiffre,
- soit un entier sans signe suivi d'un chiffre.

Un signe est :

- soit le symbole de base +,
- soit le symbole de base -.

Schéma n° 1Schéma n° 2

Un chiffre est :

- soit le symbole de base 0,
- soit le symbole de base 1,
- ... etc,
- soit le symbole de base 9.

REMARQUE

Nous attirons l'attention du lecteur sur la définition de "entier sans signe". Celle-ci est récursive, c'est-à-dire que dans la définition de cette notion (dans l'ovale), nous retrouvons la même notion (dans le rectangle). Donnons un exemple de formation d'un entier sans signe :

8 est un entier sans signe (car un entier sans signe peut être un chiffre).

82 est un entier sans signe (car c'est l'entier sans signe 8 suivi du chiffre 2).

827 est un entier sans signe (car c'est l'entier sans signe 82 suivi du chiffre 7).

B - Cartes syntaxiques du langage de description.

B - 1. Document (voir schéma n° 3, page 2.6)

La forme la plus complète d'un document est donc :

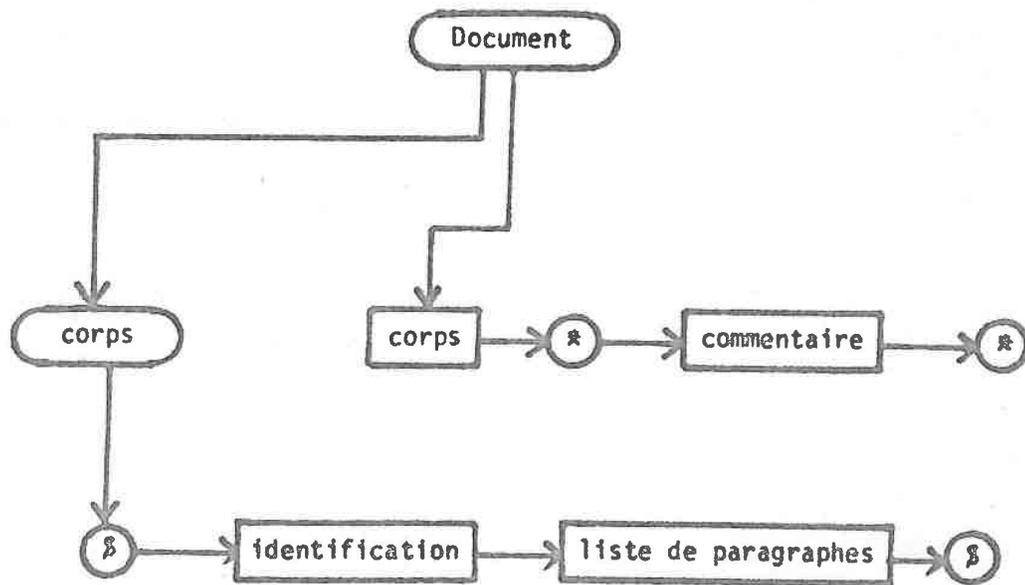
§ identification liste de paragraphes § * commentaire *

La partie essentielle d'un document et la plus intéressante est la liste de paragraphes. Toutefois, nous allons d'abord expliciter ce qu'est un commentaire (B - 2), puis l'identification (B - 3) afin de nous libérer de ces notions.

B - 2. Commentaire

Le commentaire est une chaîne de caractères alphanumériques autre que l'étoile (*) de longueur quelconque. Notons que le caractère "blanc" est significatif et sert de séparateur. Le commentaire permet à l'utilisateur de préciser le sens d'un document ou d'y porter une appréciation par exemple. Précisons que le commentaire ne pourra pas faire l'objet de questions au moment de l'interrogation. Il est tout de même intéressant de le donner car il sera restitué dans les réponses fournies.

Schéma n° 3



B - 3. Identification. (voir schéma n° 4, page 2.8)

L'entête permet de donner des précisions sur le document. Chacune d'elles doit être présente conformément à la carte syntaxique précédente.

Codeur.

Il s'agit d'un caractère numérique ou alphabétique. Celui-ci permet d'identifier la personne qui a fait la description du document. Ce renseignement sera utile au moment de la correction des documents que la machine déclare erronés (voir chapitre III).

Numéro.

Ce numéro est un nombre entier de trois chiffres exactement. Il est attribué au document par le codeur qui en a fait la description. Chaque codeur numérote ses documents et ceci par ordre croissant à partir de 1. On évite ainsi les problèmes d'attribution de numéros corrects aux documents puisque chaque codeur s'occupe de son numérotage. En outre, cela permet d'utiliser moins de caractères pour attribuer des numéros aux documents.

Référence.

La référence est un nombre entier de trois chiffres. Celui-ci désigne le nom de la revue ou du livre dont est extrait le document. Nous évitons ainsi à l'utilisateur de recopier des noms assez longs qui apparaissent fréquemment.

Année.

C'est l'année de parution de l'article donnée sur quatre chiffres,

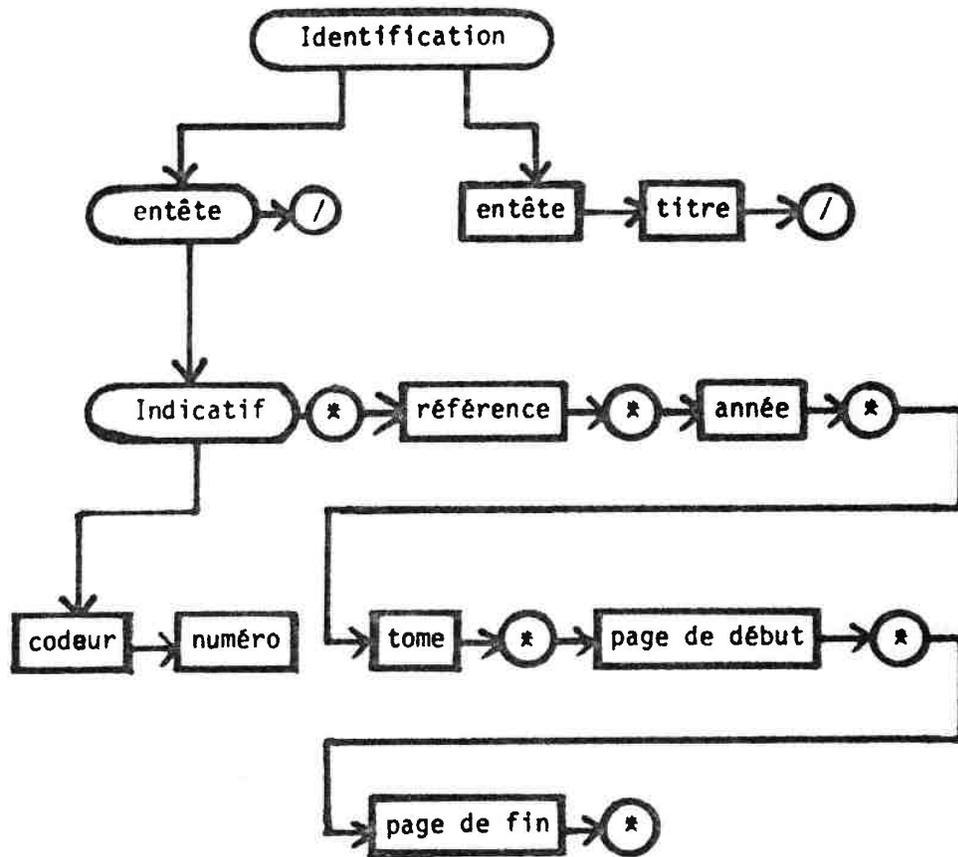
Tome.

Ce numéro de trois chiffres représente soit un numéro de revue, soit un numéro de tome suivant la provenance du document (revue ou livre).

Page de début.

C'est le numéro sur 4 caractères de la page de début du document dans la revue ou dans le livre dont il est extrait.

Schéma n° 4



Page de fin.

C'est le numéro sur 4 caractères de la page de fin du document dans la revue ou dans le livre dont il est extrait.

Titre.

Le titre est une chaîne de caractères alphanumériques de longueur quelconque qui permet de donner le titre du document. Le caractère "blanc" est significatif et peut donc servir de séparateur. Précisons que les noms d'auteurs quant à eux sont donnés dans le paragraphe de nom AUT (car leur nombre est variable et sont classés en deux catégories : auteur et co-auteur).

B - 4. Liste de paragraphes. (voir schéma n° 5, page 2.10)

Un nom de paragraphe comporte au plus cinq caractères alphabétiques ou numériques. Ces noms ont été choisis une fois pour toute et sont répertoriés dans le lexique des noms de paragraphes. Actuellement une vingtaine de termes ont été retenus, ce nombre pouvant aller sans problème jusqu'à cinquante (nombre prévu dans les programmes). Nous donnons ici quelques exemples de noms de paragraphe qui seront utilisés dans la suite de l'exposé :

AUT	auteur
BALPG	balloon pumping
PATT	patient
RESUL	result

Le choix des noms de paragraphes n'a pas été facile à faire car, comme nous l'avons mentionné dès l'introduction (II. 1), un paragraphe correspond à un découpage du document en rubriques.

Pour un document donné tous les noms de paragraphes ne sont pas nécessairement présents. Seuls figurent dans la liste de paragraphes ceux dont on a besoin pour décrire le document. Par exemple le paragraphe de nom ANEST (anesthésie) est absent d'un document qui ne parle pas d'anesthésie.

B - 5. Liste de vecteurs. (voir le schéma n° 6, page 2.10)

Nous donnons un exemple formel de liste de paragraphes où les P_i sont des noms de paragraphe et les V_i des vecteurs :

$$P1 = V1 \neq V2 \neq V3 / P2 = V4 / P3 = V5 \neq V6 /$$

Schéma n° 5

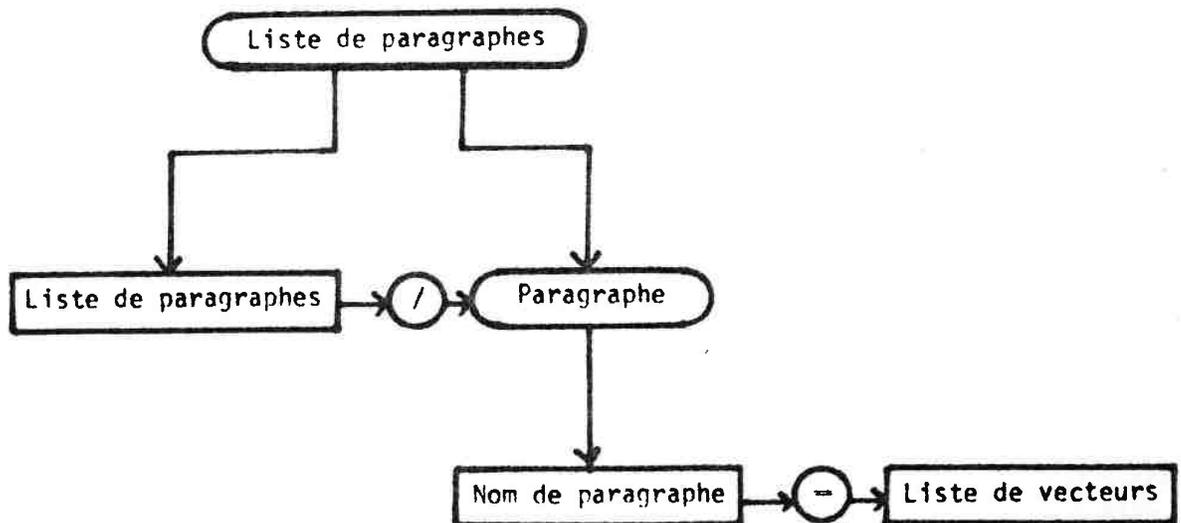
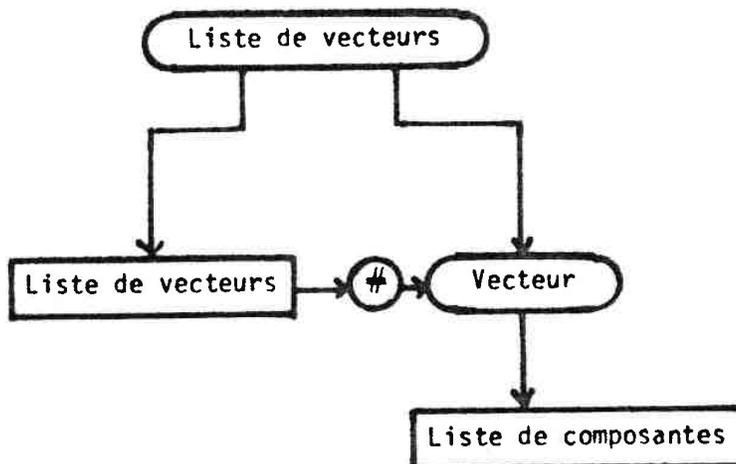


Schéma n° 6



B - 6. Liste de composantes. (Voir le schéma n° 7, page 2.12)

Un nom de composante comporte au plus cinq caractères alphabétiques ou numériques. Un nom de composante n'est en général usité que dans quelques paragraphes, voire un seul. L'utilisateur dispose, pour chaque paragraphe, de la liste des composantes qui y sont utilisables. Un autre intérêt de classer les noms de composantes par paragraphe est de pouvoir faire des contrôles au moment de l'entrée des documents en machine.

Par exemple, quelques uns des noms de composantes attachés au paragraphe des patients (PATT) sont :

SPEC	specious
NBR	number
AGE	age
WGHT	weight
SEX	sex

L'utilisateur est tenu, pour un paragraphe donné, de se servir uniquement des noms de composantes qui y sont rattachés.

B - 7. Liste de valeurs. (Voir le schéma n° 8, page 2.13)

Nous allons préciser les notions de mot, de nombre et de valeur standard.

Mot.

Un mot est une chaîne de vingt quatre caractères alphabétiques ou numériques au plus, le premier étant obligatoirement une lettre. Un mot ne peut pas être l'une des chaînes de caractères suivantes :

E, I, A, D, P, NS, PC.

Les mots que l'utilisateur peut employer pour décrire des documents sont répertoriés dans un lexique dont une copie est à sa disposition. Les mots permis y sont classés par nom de composante afin de simplifier le travail de description des documents. Notons que l'orthographe des mots donnée par le lexique doit être respectée, en particulier les mots sont toujours écrits au singulier.

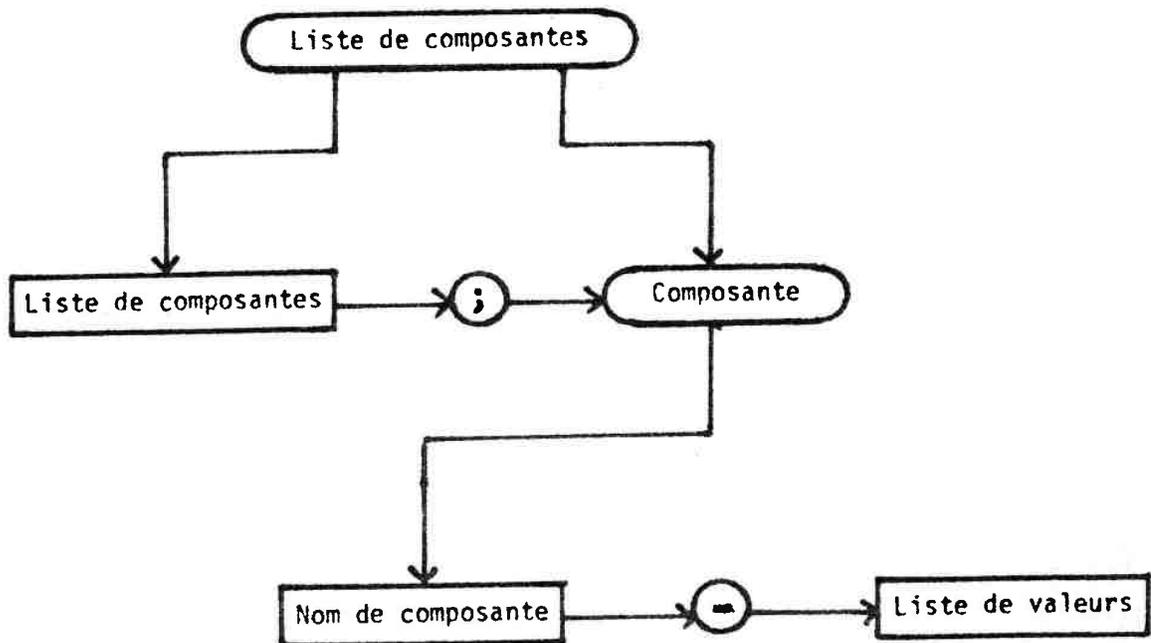
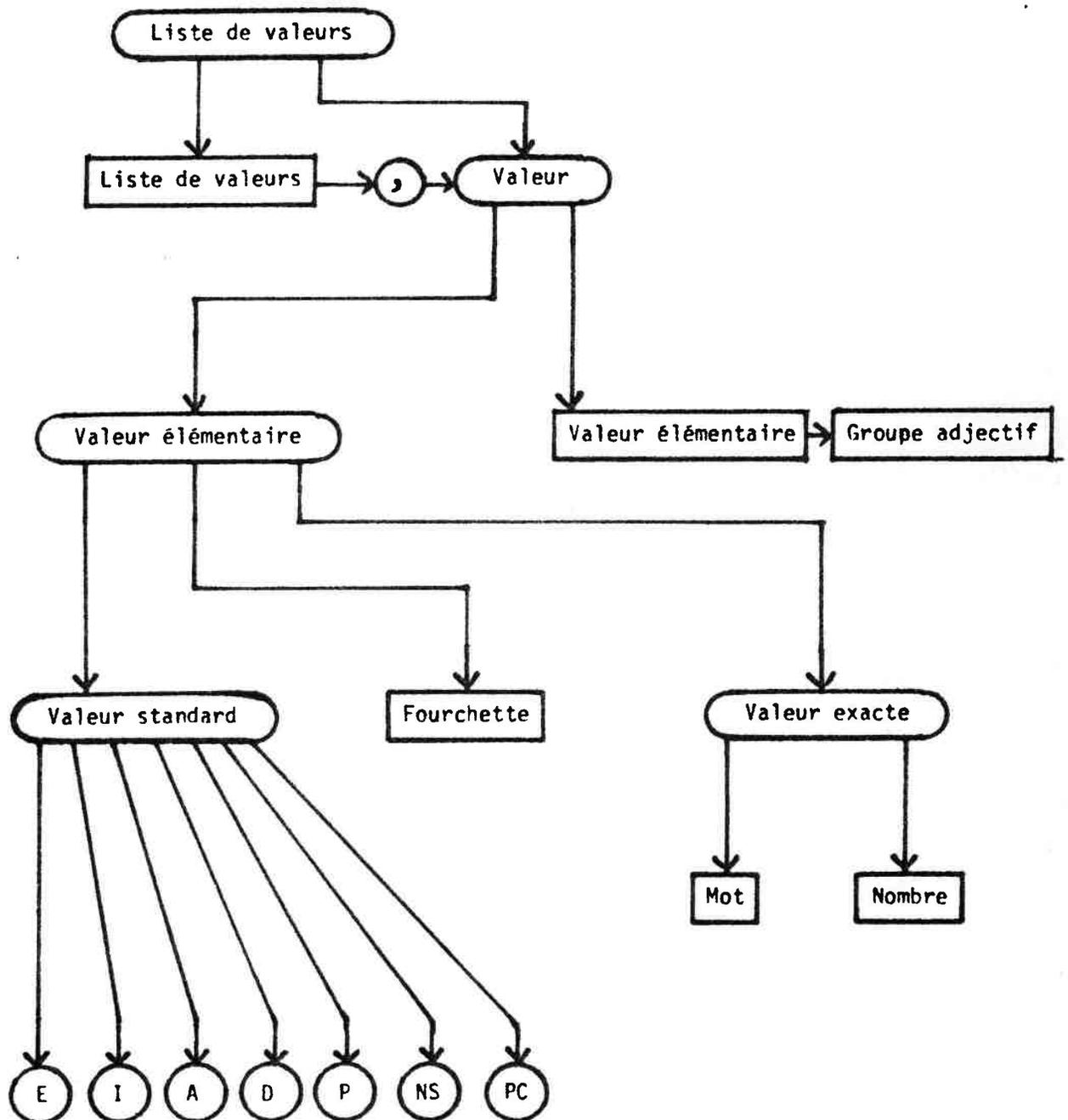
Schéma n° 7

Schéma n° 8



La construction de ce lexique nous a posé des problèmes que nous évoquerons plus loin (en III. B) au moment de sa description détaillée. Signalons que ce lexique est organisé, nous y avons en effet introduit une relation de hiérarchie entre les mots.

Nombre.

Un nombre est :

- soit un nombre entier
- soit un nombre réel

Le format d'un nombre entier est : $\pm n$

Un nombre réel a trois formats possibles : $\pm n.d$

$\pm n.d E \pm s$

$\pm n E \pm s$

n : série de neuf chiffres décimaux au plus, d'un chiffre au moins

d : série de neuf chiffres décimaux au plus

E : symbole qui désigne l'exponentiation de base 10

s : série de deux chiffres décimaux au plus qui représente l'exposant de 10.

\pm : désigne soit +, soit -. Le signe + peut être omis aussi bien devant le nombre que devant l'exposant (s).

Exemple de nombres.

$1 E 1$ vaut $1 \times 10^1 = 10$

$- 12.36 E - 1$ vaut $-(12,36 \times 0,1) = - 1,236$

Valeur standard.

E : Une valeur qui a été étudiée mais qui n'a pas pu être déterminée est décrite par la valeur E . On a par exemple essayé de doser la quantité d'un produit X sans y parvenir. On écrira la composante suivante :

$X = E$ (où X est un nom de composante)

I : Certaines valeurs peuvent ne pas être précisées dans un document. Elles sont alors décrites par I (indéfini). Il arrive par exemple qu'il y ait des cases blanches dans un tableau de nombres, celles-ci sont alors décrites par I . Cette valeur standard est souvent utilisée pour satisfaire à la règle que les composantes d'un même vecteur doivent avoir le même nombre de valeurs. Nous y reviendrons dans la partie II. 3 (Utilisation du langage de description).

Nous expliquerons dans cette partie (en II. 3 - D) l'utilisation des autres valeurs standard.

Il nous reste à étudier les fourchettes pour terminer l'étude des valeurs élémentaires (voir schéma n° 8, page 2.13).

Fourchette.

(Voir le schéma n° 9, page 2.16)

Les exemples suivants montrent les trois types d'utilisation des fourchettes.

- < 1.2 : 4.3 > désigne une valeur comprise entre 1,2 et 4,3 bornes incluses.
- > 4 E 2 désigne une valeur supérieure ou égale à 400.
- < 36 désigne une valeur inférieure ou égale à 36.

Remarquons qu'une borne est toujours incluse dans une fourchette. Ce choix a été fait car les informations se présentent ainsi dans les documents médicaux. D'autre part une fourchette du type > x désigne dans l'esprit des médecins une valeur supérieure à x mais relativement proche de x. Même remarque pour les fourchettes de type < x.

Pour terminer l'étude de la syntaxe de notre langage, il reste à définir les groupes adjectif.

P - 8. Groupe adjectif. (Voir schéma n° 10, page 2.17).

Une valeur élémentaire suivie d'un groupe adjectif sera dite adjectivée. Les adjectifs de niveau 1 se rapportent à la valeur élémentaire adjectivée. Un adjectif de niveau 1 peut lui-même être précisé par une liste d'adjectifs de niveau 2.

Exemples.

Les prénoms d'auteurs sont donnés à l'aide d'adjectifs de niveau 1.

Les noms de composants MAIN (principal) et COWR (co-worker) concernent respectivement les auteurs principaux et les co-auteurs.

AUT = MAIN = DORMANDY (JA) ~~##~~

COWR = GOETZ (RH), KRIPKE (JC) /

Schéma n° 9

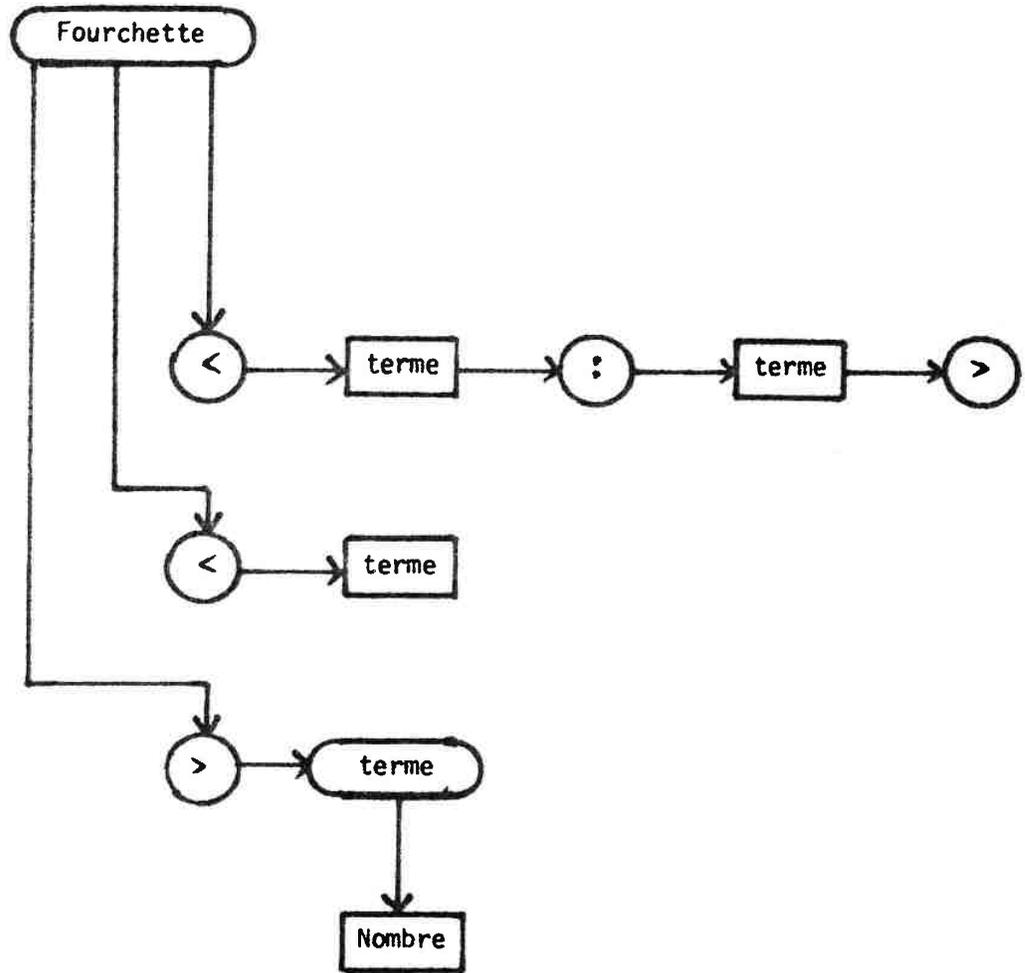
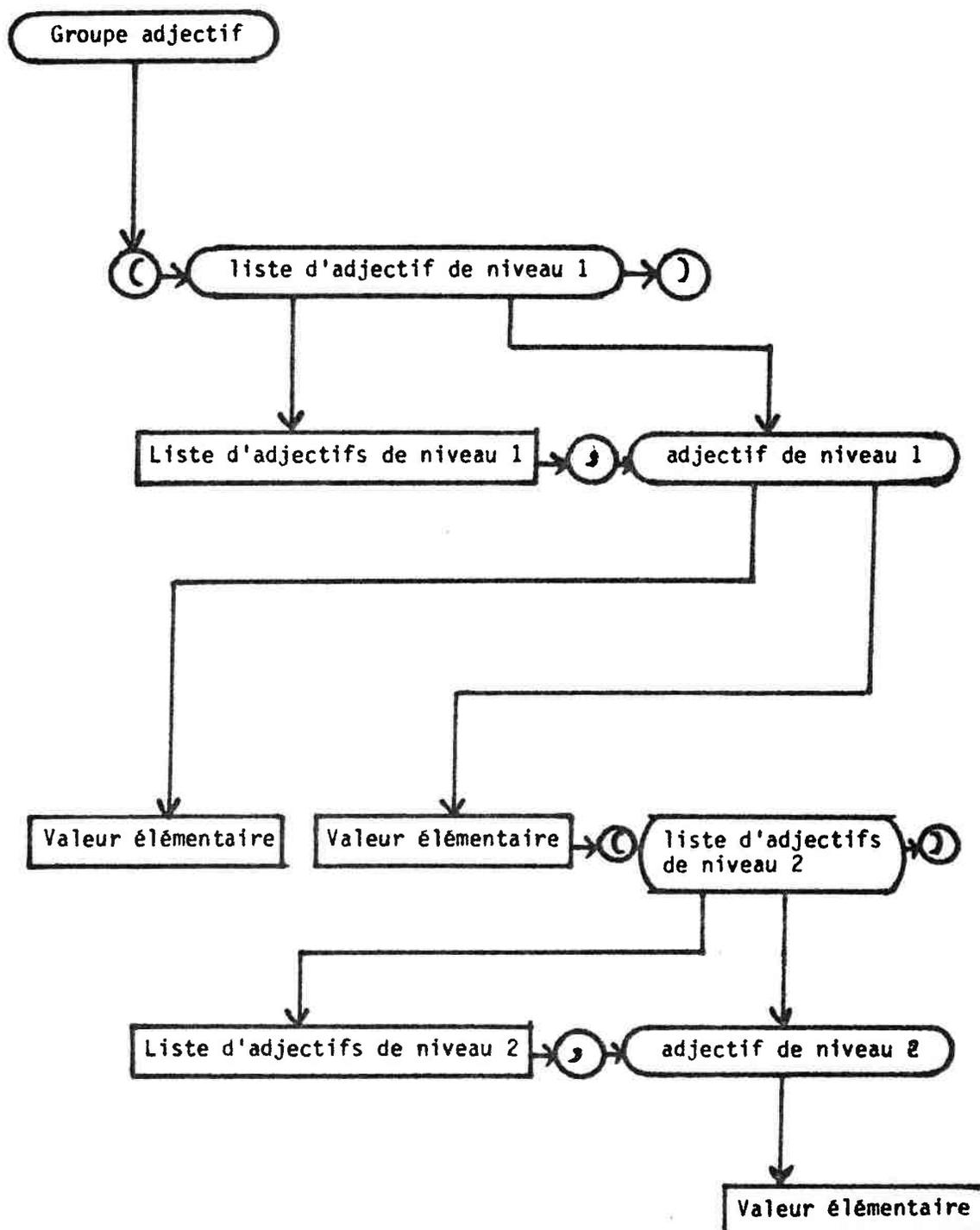


Schéma n° 10



Autre exemple :

PATT = SPEC = DOG (MONGREL, GREY (VERY)) ;
 WGT = 15 ;
 NBR = 4 /

Le vecteur du paragraphe de nom PATT (patient) ci-dessus décrit
 4 chiens bâtards d'un gris soutenu et de 15 kg.

Autre exemple :

MYOC - INFARCT (ACUTE)

Nous donnons en II. 3 - D tout ce qui concerne l'utilisation de
 l'adjectivation (point de vue sémantique).

Remarques.

1) Une fourchette peut être adjectivée, mais non pas ses termes :

< x : y > (z, w (h, j)) est correct

< x : y (j) > est incorrect

< y (z) est correct. L'adjectif z concerne la fourchette
 < y et non pas le terme y.

2) Une fourchette peut être utilisée en tant qu'adjectif de niveau

1 ou 2

x (< y : z >)

x (y (< z))

II. 3 - UTILISATION DU LANGAGE DE DESCRIPTION.

Après l'étude de la syntaxe du langage (II. 2) nous abordons ici la manière de s'en servir.

A - Description des expériences.

Dans un même document, il y a en général l'étude de plusieurs expériences sur des patients différents, avec des appareillages différents... etc. Toutes ces informations sont décrites par des vecteurs eux-mêmes ventilés dans des paragraphes. Les informations qui concernent une même expérience sont donc disséminées dans la description du document. Il est donc nécessaire d'avoir à sa disposition un moyen de désigner les vecteurs d'une même expérience.

Dans un premier temps, nous définissions les différentes expériences par l'énumération des vecteurs qui la composent. Cette méthode s'est rapidement révélée lourde et surtout fastidieuse à mettre en oeuvre pour l'utilisateur. D'où la recherche d'une autre solution.

Nous avons alors pensé que la plus simple serait d'imiter ce que fait en réalité une personne organisée qui prend des notes concernant une expérience.

Les expériences sont numérotées et les numéros ainsi définis sont attachés aux informations qui concernent chacune d'entre elles. Prenons un exemple.

Exemple.

Un document décrit deux expériences.

1. Sur des chiens de 15 à 20 kg ayant subi un infarctus du myocarde expérimental, il a été réalisé une assistance cardio-circulatoire au moyen d'un ballonnet intra-aortique de volume 20 cc et de longueur 28 cm à une seule chambre. On a mesuré la pression aortique moyenne (entre 100 et 120 mm Hg) et la pression artérielle moyenne en oxygène (120 mm Hg).

2. Des chiens de plus de 25 kg dans les mêmes conditions (infarctus plus assistance) et de fréquence cardiaque supérieure à 110, ont donné les mêmes résultats que ceux de la première expérience. En outre le débit cardiaque est de 1,1 l/minute.

Dans le schéma n° 11 (page 2.21), les rectangles représentent les informations "qui vont ensemble". Ce sont pour nous des vecteurs. Les numéros à droite de ces rectangles désignent leur appartenance à l'expérience n° 1 ou n° 2. Quant aux accolades, elles regroupent les vecteurs en paragraphes.

A partir de ce schéma, nous allons décrire ce document à l'aide de notre langage. Pour chaque vecteur, nous précisons son appartenance à une expérience à l'aide de la composante de nom EXP. Les autres noms de composante utilisés sont :

SPEC	specious
WGHT	weight
BALCA	balloon capacity
LENG	length
CHAMB	chamber
MAOP	Mean aortic pressure
APO2	arterial partial pressure O ₂
CO	cardiac output
COND	condition
HRATE	heart rate

Les noms de paragraphe utilisés sont :

PATT	patient
BALPG	balloon pumping
RESUL	result
PROC	procedure

PATT = SPEC = DOG ; WGHT = < 15 : 20 > ; EXP = 1 *

SPEC = DOG ; WGHT = > 25 ; EXP = 2 /

Patients	chiens de 15 à 20 kg	1
	chiens de plus de 25 kg	2
Pompage par Ballonnet	ballonnet volume = 20 longueur = 28 chambre = 1	1 et 2
Résultats	pression aortique moyenne entre 100 et 120	1 et 2
	pression en oxygène 120	1
	débit cardiaque 1,1	2
Procédure	condition : infarctus expérimental	1 et 2
	fréquence cardiaque plus de 110	2

BALPG = BALCA = 20 ; LENG = 28 ; CHAMB = 1 ; EXP = 1,2 /

PROC = COND = MYOC - INFARCT (EXPERIMENTAL) ;

EXP = 1,2 ≠

HRATE = > 110 ; EXP = 2 /

RESUL = MAOP = < 100 : 120 > ; EXP = 1,2 ≠

CO = 1.1 ; EXP = 2 ≠

APO2 = 120 ; EXP = 1 /

Rappelons que l'ordre d'écriture des composantes dans un vecteur n'a pas d'importance, EXP peut donc figurer à n'importe quelle place (dans le vecteur concerné).

L'écriture EXP = 1, 2 signifie : pour les expériences 1 et 2. La liste de valeurs de la composante EXP est obligatoirement constituée de numéros (ceux-ci peuvent aller de 1 à 99).

Un même vecteur peut donc appartenir à plusieurs expériences. Cette commodité n'est pas de pure forme car il arrive souvent qu'un vecteur soit commun à plusieurs expériences et par conséquent cela évite de recopier plusieurs fois la même chose (et les vecteurs peuvent être longs).

On remarquera aussi que plusieurs vecteurs d'un même paragraphe peuvent appartenir à la même expérience (deux des trois vecteurs du paragraphe RESUL concernent l'expérience n° 1 dans l'exemple).

Cet exemple sert d'introduction à la notion d'expérience (au sens où nous l'entendons).

Notion d'expérience.

Une expérience est décrite par un ensemble de vecteurs qui se distingue d'un autre ensemble par un vecteur au moins.

Il y a des expériences (au sens où nous l'entendons) qui n'apparaissent pas a priori à la lecture d'un document.

Exemple :

Dix chiens mâles de race Beagle de 10 à 15 kg, de plus de 3 ans d'âge ont subi la même expérimentation. Trois d'entre eux ont donné tel résultat, sept ont donné tel autre résultat.

Nous sommes amenés dans un tel cas à décrire deux expériences (au sens défini ci-dessus).

```
PATT= SPEC = DOG (BEAGLE) ; WIGHT = < 10 : 15 > ;
      NBR = 3 ; AGE = > 3 ; SEX = M ; EXP = 1 #
      SPEC = DOG (BEAGLE) ; WIGHT = < 10 : 15 > ;
      NBR = 7 ; AGE = > 3 ; SEX = M ; EXP = 2 /
```

Description de l'expérimentation

```
RESUL = EXP = 1 ;----- # EXP = 2 ; ----- /
```

Les deux vecteurs du paragraphe PATT ont toutes leurs composantes identiques sauf NBR et EXP. Nous verrons plus loin (II. 3 - E) que l'écriture de ce paragraphe peut être notablement allégée en utilisant une facilité offerte par notre langage lorsque deux vecteurs ont beaucoup de renseignements en commun.

Il découle de la façon de décrire les expériences qu'un vecteur qui ne contient pas la composante EXP n'est attaché à aucune expérience. En d'autres termes, toutes les expériences sont concernées par un tel vecteur. Généralement la donnée des noms d'auteurs par exemple concerne tout le document, et donc chaque expérience (Il est d'ailleurs possible d'attacher des noms d'auteurs à une expérience en utilisant la composante EXP, si besoin est).

Nous allons donner un résumé de ce qui concerne l'utilisation de la composante EXP.

Résumé.

1. EXP peut être utilisée dans n'importe quel paragraphe.
2. Les valeurs de la liste de valeurs de EXP sont obligatoirement des nombres entiers entre 1 et 99, bornes incluses. Le nombre des valeurs de la liste n'est pas limité (un même vecteur peut appartenir à plusieurs expériences).

3. L'utilisateur est libre de numéroté les expériences comme il l'entend (pas nécessairement à partir de 1 et par ordre croissant comme dans les exemples).
4. EXP fait exception à la règle déjà signalée que les composantes d'un même vecteur doivent avoir le même nombre de valeurs.
5. Plusieurs vecteurs d'un même paragraphe peuvent concerner la même expérience
6. Un vecteur qui ne contient pas EXP concerne toutes les expériences.

B - Description des tableaux.

Il arrive souvent que des informations soient données sous forme de tableau dans les documents. Notre langage permet de les décrire. Mais prenons d'abord un exemple. Le tableau suivant donne l'évolution de l'hémoglobine du plasma et du nombre de globules rouges en fonction du temps.

Temps	1	2	3	
Hémoglobine	40	80	102	(mg/100ml)
Globules rouges	4,2	4	3,5	(million/mm ³)

Les instants sont repérés par les numéros 1, 2 et 3.

Dans notre langage, ce tableau est décrit à l'aide d'un vecteur, dont chaque composante représente une ligne. Les noms de composantes à utiliser sont :

TIME temps
 PHB plasmatic hemoglobin
 RBC red blood cells

d'où la description du tableau :

TIME = 1, 2, 3 ;
 PHB = 40, 80, 102 ;
 RBC = 4.2, 4, 3.5 #

Les valeurs des composantes doivent être décrites dans l'ordre où elles apparaissent dans les lignes du tableau. En effet, les valeurs de même rang des différentes composantes sont en relation. En conséquence, toutes les listes de valeurs doivent avoir le même nombre d'éléments. Des cases "blanches" peuvent exister dans un tableau. Elles sont remplacées dans la description par la valeur standard I (voir II. 2 - B - 6).

Si l'on veut exprimer que l'on a obtenu les valeurs 80, 110, 102 pour la fréquence cardiaque et les valeurs 52, 48, 45 pour l'hématocrite sans qu'il n'y ait de relation entre les deux listes de valeurs, on fabriquera deux vecteurs :

HRATE = 80, 110, 102 # HTE = 52, 48, 45 #
(HRATE signifie heart rate et HTE hematocrit)

En d'autres termes, un vecteur formé de plusieurs composantes décrit en réalité un tableau. Le vecteur suivant :

SPEC = DOG ; WGHT = < 18 : 20 > ; AGE = 4 #
peut être considéré comme la description du tableau :

SPEC	DOG
WGHT	<18 : 20>
AGE	4

Il existe bien une relation entre les valeurs de même rang des différentes composantes.

Pour indiquer l'appartenance d'un tableau à une expérience, il suffit d'utiliser la composante EXP. Pour exprimer que le tableau donné en exemple ci-dessus fait partie de l'expérience n° 7 on écrira :

TIME = 1 , 2 , 3 ;
PHB = 40 , 80 , 102 ;
RBC = 4.2 , 4 , 3.5 ;
EXP = 7 #

La composante EXP fait exception à la règle qui veut que les composantes d'un même vecteur doivent avoir le même nombre de valeurs.

Il convient d'ailleurs de ne pas interpréter l'écriture :

NA = 140, 145 ;

K = 3.8, 4.5 ;

EXP= 3 , 9 #

comme étant la traduction du tableau :

NA	140	145
K	3,8	4,3
EXP	3	9

Le vecteur ci-dessus signifie en effet que le tableau ci-dessous est commun aux expériences n° 3 et n° 9 :

NA	140	145
K	3,8	4,3

Remarquons que dans un même tableau, les éléments peuvent être n'importe quelle valeur, c'est-à-dire des mots, des nombres, des fourchettes, des leurs standard, avec adjectivation ou pas.

Nous donnons un résumé de ce qui concerne la description des tableaux .

Résumé.

1. Un tableau est décrit par un vecteur dont chaque composante correspond à une ligne (ou bien à une colonne si le tableau est présenté différemment).

2. Il existe des relations entre les valeurs de même rang des composantes d'un même vecteur (relation implicite).
3. Les composantes d'un même vecteur doivent avoir le même nombre de valeurs, La valeur standard I permet de satisfaire à cette règle si le tableau possède des cases "blanches".
4. La composante EXP fait exception aux points 2 et 3.

C - Description de l'échelle des temps.

C - 1. Schéma de l'échelle des temps.

Nous abordons ici le problème qui est de définir la chronologie des événements qui sont survenus pendant le déroulement d'une expérimentation.

Les instants sont souvent définis par rapport à des événements aléatoires. Par exemple : telle injection a été faite cinq minutes après le début de l'infarctus du myocarde, ou encore : la pression partielle artérielle en oxygène était de 150 mm Hg 10 minutes avant la mort du patient.

D'un document à l'autre, voire à l'intérieur d'un même document, les repères qui concernent le temps n'ont rien de comparables. Par conséquent il serait impossible d'interroger les documents sans l'établissement d'un système général de représentation de l'échelle des temps.

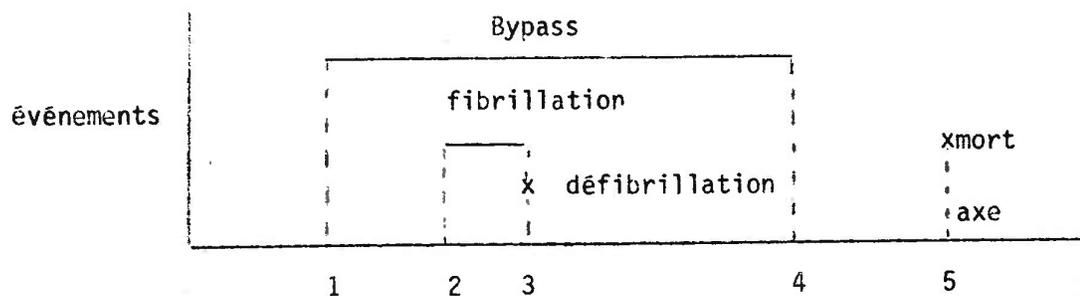
Nous avons été amenés à représenter la suite des événements concernant une expérience à l'aide d'un schéma (qui est ensuite traduit à l'aide de notre langage de description). Notons que ce schéma n'est pas indispensable pour coder un document. Il offre cependant une méthode de travail à l'utilisateur (bien utile pour certains documents où les informations concernant la chronologie sont très disséminées).

Les événements sont représentés au dessus d'un axe des temps orienté de gauche à droite. Les événements ponctuels sont matérialisés par une croix, ceux qui durent sont représentés par un segment de droite. En dessinant sur plusieurs niveaux, il est facile de représenter des événements qui se chevauchent. Les extrémités des segments et les croix sont projetées sur l'axe. Les repères ainsi obtenus sont numérotés à partir de 1 et de gauche à droite.

Exemple.

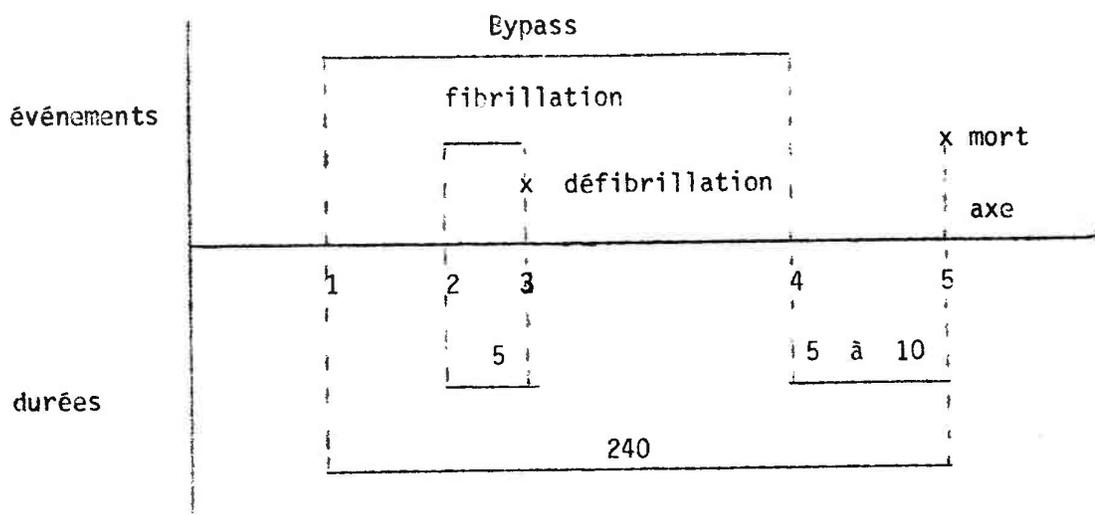
Pendant le bypass effectué dans une expérimentation, il y a eu une fibrillation ventriculaire qui a duré cinq minutes. Elle s'est terminée par une défibrillation par choc électrique. Le bypass a été arrêté entre cinq et dix minutes avant la mort du patient qui est survenue quatre heures après le début du bypass.

La représentation des événements est la suivante :



Lorsque le texte précise certaines durées, celles-ci sont représentées par des segments de droite que l'on dessine en dessous de l'axe des temps. Ces durées sont exprimées en minutes.

Complétons le schéma précédent :



Un tel schéma permet donc de replacer les événements les uns par rapport aux autres et de préciser les durées qui sont connues.

C - 2. Description du schéma de l'échelle des temps.

Un tel schéma peut se résumer à l'aide de deux tableaux. Un premier tableau définit les repères :

Repère	de 1 à 4	de 2 à 3	5	3
Événement	Bypass	Fibrillation	Mort	Défibrillation

Un deuxième tableau définit les durées connues exprimées en minutes en fonction des repères :

Laps	de 1 à 5	de 2 à 3	de 4 à 5
Durée	240	5	de 5 à 10

Nous savons décrire des tableaux (voir II. 3 - B). Les composantes à utiliser sont :

REP repère
 EVENT event
 LAPS lapse
 DUR duration

Les vecteurs qui décrivent ces tableaux sont ventilés dans le paragraphe de nom CHRON (chronology)

CHRON = REP = <1 : 4 >, <2:3> , 5, 3 ;
 EVENT = BYPASS, FIBRILL (VENTRICULAR), DEATH,
 DEFIBRILL (SHOCK (ELECTRIC)) #
 LAPS = <1:5> , <2:3> , <4:5> ;
 DUR = 240 , 5 , <5:10> /

Si dans un vecteur d'un autre paragraphe, on veut donner une indication concernant le temps, celle-ci doit obligatoirement faire référence aux repères définis dans le paragraphe CHRON.

Pour exprimer que le sodium était de 136,1 au début du bypass, on écrira le vecteur :

$$NA = 136,1 ; TIME = 1 \neq$$

Pour exprimer que le sodium était de 138 à 140 pendant la fibrillation, on écrira :

$$NA = \langle 138:140 \rangle, TIME = \langle 2:3 \rangle \neq$$

Les valeurs de la liste de valeurs de la composante TIME sont obligatoirement :

- soit des repères
- soit des fourchettes dont les termes sont des repères,

Ces valeurs peuvent être adjectivées par un nombre ou une fourchette. Les adjectifs (de niveau 1) expriment un décalage en minutes par rapport à la valeur adjectivée.

Donnons des exemples (les repères sont ceux définis ci-dessus).

TIME = 2(3) : 3 minutes après le début de la fibrillation.
 TIME = 3(-10) : 10 minutes avant la fin de la défibrillation.
 TIME = 3(<5:10>) : 5 à 10 minutes après la défibrillation.
 TIME = 1(<-5:10>) : 5 minutes avant à 10 minutes après le début du bypass.

C - 3. Remarques.

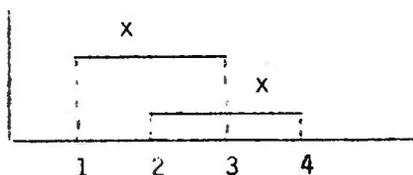
Remarque 1.

Si plusieurs événements ont même extrémité, celle-ci ne détermine qu'un seul repère. Dans l'exemple, la fin de la fibrillation et la défibrillation correspondent au seul repère numéro 3.

Remarque 2.

Le système de repérage utilisé permet de définir sans ambiguïté les extrémités de deux événements de même nom qui se chevauchent.

Événements



Ce cas ne s'est pas présenté jusqu'à présent dans les documents étudiés. D'où l'événement de nom X dans le schéma ci-dessus.

Remarque 3.

Lorsque plusieurs expériences d'un même document ont des chronologies différentes, c'est-à-dire que les repères des expériences ne sont pas communs, il suffit de les décrire toutes en précisant dans les vecteurs obtenus (du paragraphe CHRON) les expériences concernées (à l'aide de la composante EXP).

Il y a alors lieu de veiller à la bonne utilisation des repères. Un même numéro peut en effet représenter des événements différents suivant l'expérience concernée.

D - Utilisation des adjectifs.

Pour que l'inventaire soit complet, nous citons les utilisations déjà signalées.

D - 1. Qualifier un mot.

Exemples : MYOCARD-INFARCT (ACUTE)
SZTUKA (JC)

D - 2. Donner un décalage par rapport à un repère de l'échelle des temps.

Le lecteur se reportera au paragraphe précédent (II. 3 - C).

D - 3. Préciser si la variation entre deux valeurs est significative ou pas.

La variation entre deux valeurs peut être due en partie au hasard. Si la probabilité que la variation entre deux valeurs soit due au hasard est grande, on dit que cette variation est non significative. Dans les documents médicaux, ces indications existent sous deux formes :

- la mention NS (non significatif)
- la donnée de la probabilité précitée.

Dans notre langage, nous procéderons comme suit :

a) Pour indiquer que la variation entre deux valeurs n'est pas significative, on fait suivre la deuxième par la valeur standard NS. (adjectivation)

Exemple : MAOP = 98, 101 (NS), 110

La variation entre 98 et 101 n'est pas significative. Les deux valeurs doivent bien sûr appartenir à la même liste de valeurs et être consécutives.

b) Pour donner la probabilité pour que la variation soit due au hasard, on utilise l'adjectif P (valeur standard). Celui-ci est adjectivé au deuxième niveau par la valeur de cette probabilité (donnée en général par une fourchette).

Exemple : MAOP = 95, 105 (P (<0.05))

La probabilité pour que la variation entre 95 et 105 soit due au hasard est inférieure à 0,05. A remarquer que cette écriture est quasiment identique à celle trouvée dans le texte des documents où les auteurs écrivent : (P<0,05).

c) Le nombre d'adjectifs n'étant pas limité, les deux indications ci-dessus peuvent être combinées :

Exemple : MAOP = 98, 103 (NS, P (>0.05))

D - 4. Décrire une valeur par ses variations.

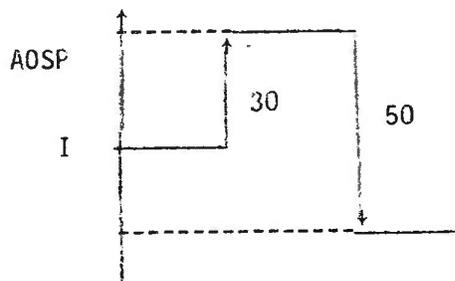
Il arrive que la donnée d'une valeur soit faite sous la forme d'une variation (augmentation ou diminution) par rapport à une autre valeur. Par exemple : la pression systolique dans l'aorte a augmenté de 30 mm Hg par rapport à une valeur initiale non chiffrée. Puis elle a diminuée de 50 mm Hg par rapport à la dernière valeur relevée.

Nous avons trois valeurs à décrire :

- la valeur initiale
- la valeur obtenue après l'augmentation
- la valeur obtenue après la diminution.

Nous pouvons représenter ces valeurs sur un schéma :

Valeurs de l'AOSP (Aortic Systolic Pressure)



Nous nous proposons de montrer sur cet exemple l'influence de l'interrogation sur les choix des méthodes de description. Dans un premier temps, nous voulions donner les valeurs de l'AOSP de l'exemple par :

$$\text{AOSP} = \text{I}, 30 \text{ (A)}, 50 \text{ (D)}$$

(AOSP est un nom de composante)

Les valeurs standards A et D (voir schéma n° 8, page 2.13) sont utilisées en tant qu'adjectifs de niveau 1 et précisent que 30 et 50 ne sont pas des valeurs de l'AOSP mais le résultat d'une augmentation (A) puis d'une diminution (D). Cette solution est mauvaise du point de vue de l'interrogation. Imaginons qu'un utilisateur demande la sélection des documents dans lesquels il existe AOSP = 120. Le programme d'interrogation doit alors systématiquement vérifier chaque fois qu'il trouve 120 pour AOSP s'il ne s'agit pas d'une variation de 120. Et ceci devra être fait pour toutes les demandes de ce type (sélectionner une valeur). D'où une perte de temps.

Nous avons retenu une autre solution :

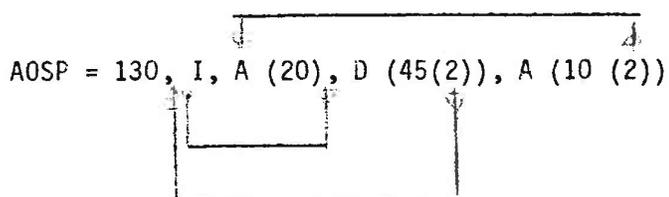
$$\text{AOSP} = \text{I}, \text{A (30)}, \text{D (50)}$$

Une valeur qui est le résultat d'une variation est décrite par A ou D, ce qui évite les inconvénients cités ci-dessus. La valeur de la variation est donnée par un adjectif de niveau 1.

Dans l'exemple précédent, les variations sont données par rapport à la valeur précédente de la liste de valeurs. Mais une variation peut être donnée par rapport à une autre valeur que celle qui la précède immédiatement dans la liste de valeurs.

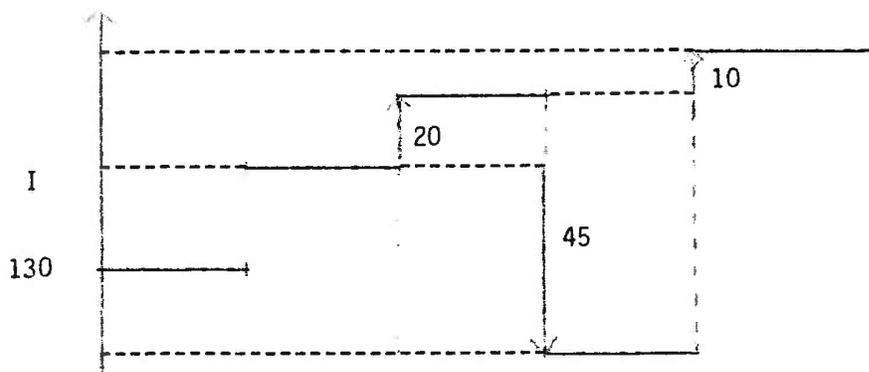
Si une variation est donnée par rapport à la $i^{\text{ème}}$ valeur qui la précède (dans la liste) sa valeur est adjectivée par i (au deuxième niveau). Par défaut de cet adjectif de niveau 2, i vaut 1.

Exemple.



D(45(2)) : valeur résultante de la diminution de 45 mm Hg par rapport à la deuxième valeur qui précède, c'est-à-dire I.

Le schéma suivant représente les valeurs de AOSP :



A remarquer que l'on peut décrire des valeurs résultant de variations non chiffrées :

AOSP = 130 , A , D (I (2))

La deuxième valeur est le résultat d'une augmentation par rapport à 130. La troisième valeur est le résultat d'une diminution par rapport à 130. Il a été nécessaire d'utiliser I afin de pouvoir préciser par rapport à quelle valeur se situe la diminution (car cette information est nécessairement un adjectif de niveau 2).

Remarques.

1. Les valeurs des variations peuvent être des fourchettes :

AOSP = I, A (<15:20>), D (<50:70>(2))

La deuxième valeur représente le résultat d'une augmentation de 15 à 20 mm Hg par rapport à la première valeur (I). La troisième valeur représente le résultat d'une diminution de 50 à 70 par rapport à la première valeur (I).

2. Le programme qui prend en compte les documents et les range en machine fait les calculs possibles à effectuer. Si la description de l'utilisateur contient :

AOSP = 135 , A (5) , D (<10:20>(2))

cette composante sera en réalité rangée sous la forme :

AOSP = 135 , 140 , <115:125>

Cela nécessite moins de place de rangement, mais surtout facilitera la recherche d'une valeur au moment de l'interrogation dans la mesure où si les valeurs standards A ou D sont rencontrées, cela signifie que les variations correspondantes ne sont pas calculables.

D - 5. Indiquer qu'une valeur est un pourcentage.

Nous avons vu que les valeurs quantitatives sont exprimées dans une unité propre à chaque composante et choisie une fois pour toute. Il arrive qu'une valeur représente un pourcentage d'une autre valeur de la même liste de valeurs.

Dans ce cas, on utilisera la valeur standard PC que l'on adjectivera au premier niveau par la valeur numérique du pourcentage. Pour indiquer par rapport à quelle valeur ce pourcentage est donné, on procédera exactement de la même manière que pour les variations (voir paragraphe précédent).

Exemple. AOSP = I, PC (80), PC (105 (2))

La deuxième valeur de la liste représente 80 % de la première valeur, la troisième 105 %.

Remarques.

Nous faisons des remarques analogues à celles qui terminent le paragraphe précédent :

1. Une valeur de pourcentage peut être une fourchette :

$$\text{AOSP} = \text{I} , \text{PC} (<80:90>)$$

2. La composante suivante :

$$\text{AOSP} = 150 , \text{PC} (<80:90>)$$

est rangée en machine sous la forme ci-dessous :

$$\text{AOSP} = 150 , <120:135>$$

D - 6. Combinaisons d'adjectifs.

Il est possible pour l'utilisateur de combiner plusieurs adjectifs pour une même valeur (puisque les listes d'adjectifs ne sont pas limitées en longueur).

Exemple. $\text{AOSP} = \text{I} , 130 , \text{A} (10 (2), \text{P} (<0.05))$

La troisième valeur est une augmentation de 10 par rapport à la valeur initiale. La variation entre la valeur 130 et la dernière valeur de la liste a une probabilité inférieure à 0,05 d'être due au hasard.

E - La composante MODIF.

Cette composante a été introduite pour apporter des commodités d'écriture à l'utilisateur dans la mesure où elle permet d'éviter d'écrire plusieurs fois la même composante dans des vecteurs différents.

Précisons que dans un même paragraphe, les vecteurs sont implicitement numérotés à partir de 1 (dans leur ordre d'écriture).

Plaçons nous dans le vecteur n° j d'un paragraphe. Si dans ce vecteur, nous écrivons la composante :

$$\text{MODIF} = n \text{ avec } n < j$$

la tâche suivante est entreprise automatiquement au moment de l'entrée des documents en machine :

- toutes les composantes du vecteur n° n sont recopiées dans le vecteur n° j à l'exception de celles dont le nom existe dans le vecteur n° j.

Reprenons le paragraphe PATT d'un exemple donné en II. 3 - A.

```
PATT = SPEC = DOG (BEAGLE) ; WGHT = <10:15>;
      NBR = 3 ; SEX = M ; AGE = 3 ; EXP = 1 #
      SPEC = DOG (BEAGLE) ; WGHT = <10:15> ;
      NBR = 7 ; SEX = M ; AGE = 3 ; EXP = 2 /
```

Nous avons signalé que cette écriture pouvait être simplifiée :
il suffit d'écrire :

```
PATT = SPEC = DOG (BEAGLE) ; WGHT = <10:15> ;
      NBR = 3 ; SEX = M ; AGE = 3 ; EXP = 1 #
      MODIF = 1 ; NBR = 7 ; EXP = 2 /
```

La composante MODIF = 1 a pour effet de recopier les composantes SPEC, WGHT, SEX et AGE du premier vecteur dans le deuxième conformément aux explications données ci-dessus. D'où une écriture moins lourde.

Afin qu'il soit plus facile à lire, nous donnons un exemple formel d'utilisation de MODIF :

- P1 est un nom de paragraphe
- les lettres majuscules sont des noms de composantes
- les lettres minuscules sont des listes de valeurs.

Nous voulons décrire trois vecteurs, chacun attaché à un numéro d'expérience :

```
Pour l'expérience n° 3 : A = a
                        B = b
                        C = c
                        D = d
                        E = e
                        F = f
                        G = g
```

Pour l'expérience n° 7, nous avons à décrire la même chose mais avec F = k.

Pour l'expérience n° 4, nous avons à décrire la même chose que pour l'expérience n° 7, mais en ajoutant $H = h$.

En utilisant MODIF, on obtient la description suivante :

$P1 = A = a ; B = b ; C = c ; D = d ; E = e ; F = f ; G = g ; EXP = 3 \neq$
 $MODIF = 1 ; F = k ; EXP = 7 \neq$
 $H = h ; MODIF = 2 ; EXP = 4 /$

MODIF = 1 du vecteur n° 2 a pour effet de recopier toutes les composantes du premier vecteur, sauf F et EXP. Le vecteur n° 2 est donc équivalent à :

A = a	
B = b	
C = c	
D = d	→ provient du vecteur n° 1
E = e	
G = g	
F = k	
EXP = 7	

Quant au vecteur n° 3, il est équivalent à :

A = a	
B = b	
C = c	
D = d	→ provient du vecteur n° 2
E = e	
G = g	
F = k	
H = h	
EXP = 4	

On aurait d'ailleurs pu écrire le vecteur n° 3 comme suit :

MODIF = 1 ; F = k ; H = h ; EXP = 4 /

Remarques.

L'utilisation de MODIF n'est possible que si les composantes des deux vecteurs concernés ont le même nombre de valeurs. Ceci pour ne pas enfreindre la règle qui veut que les composantes d'un même vecteur aient le même nom-

-bre de valeurs. Nous avons vu que EXP faisait exception à cette règle, il en est de même pour MODIF qui a d'ailleurs obligatoirement une seule valeur.

Après l'enregistrement des documents en machine, il ne subsiste plus de trace de la composante MODIF. On obtient d'ailleurs un gain de place car seuls sont en réalité copiés les noms de composantes d'un vecteur à l'autre (les listes de valeurs ne le sont pas - voir III. C).

Nous recommandons vivement aux utilisateurs l'emploi de cette composante, si possible évidemment. Outre une écriture moins longue (et donc moins de perforations à faire) elle permet des économies de place en machine (voir alinéa précédent).

II. 4 - PRESENTATION D'UN BORDEREAU DE PERFORATION.

Les documents doivent être décrits (à l'aide de notre langage de description) sur des bordereaux dont le contenu est ensuite perforé sur carte (une ligne par carte).

A - Numérotage.

Les cartes d'un même document sont numérotées par ordre croissant à partir de zéro. Ce numéro occupe les deux premières cases de chaque ligne du bordereau. Ce numérotage permet de faire des vérifications à la lecture.

B - Partie identification.

Les différents renseignements de l'entête d'un document sont écrits sur le bordereau en format fixe, c'est-à-dire en respectant scrupuleusement le schéma n° 11, page 2.41.

L'entête peut être suivie d'un titre. Celui-ci est écrit à partir de la colonne 32 de la première carte du document. Le titre peut se prolonger sur une ou plusieurs lignes suivant sa longueur (les lignes doivent être remplies entièrement et numérotées). Rappelons que le titre se termine par un slash (/).

Si l'identification ne contient pas de titre, on met ce slash en colonne 32.

C - Partie liste de paragraphes.

Cette partie du document est écrite après le slash qui termine l'identification. L'écriture en est libre, c'est-à-dire que l'utilisateur peut présenter son document comme il l'entend (en suivant bien sûr les règles de syntaxes). Les blancs ne sont pas significatifs ce qui permet "d'aérer" l'écriture (intéressant pour les corrections). En particulier, l'utilisateur peut commencer cette partie en allant à la ligne.

BORDEREAU DE PERFORATION

I.U.C.A.

N° COMPTE: _____ DATE DÉPÔT: _____ PAGE: 1

DISPOSITION DE L'ENTÊTE ET EXEMPLE DE PRÉSENTATION.

3	S	8	11	15	17	20	21	25	26	30	31	35	40	45	50	55	60	65	70	72	75	EC
00\$	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*	↑	*
	INDICATIF	↑	REFERENCE	↑	ANNÉE	↑	TOME	↑	PAGE DEBUT	↑	PAGE FIN	↑	TITRE /									

NUMERO DE SEQUENCE

Schéma n° 11

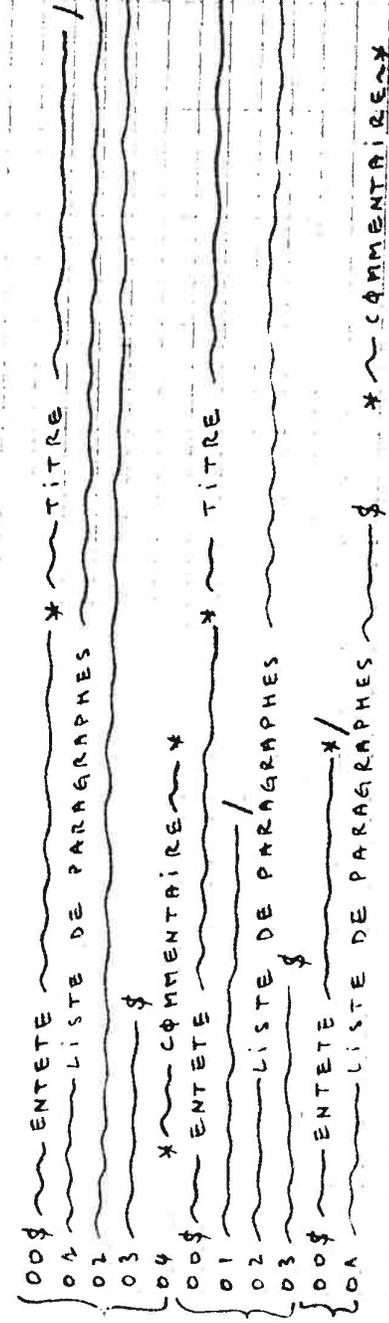


Schéma n° 12

D - Partie commentaire.

Rappelons que le corps d'un document se termine par §. S'il n'y a pas de commentaire, la description du document se termine là.

S'il y a un commentaire, celui-ci est écrit après le § qui signale la fin du corps. Il est délimité par deux étoiles. Le nombre de blancs entre § et la première étoile est indifférent (on peut aller à la ligne). Entre les étoiles, les blancs sont significatifs. Dans ce cas, le document se termine sur la deuxième étoile.

Nous donnons sur le schéma n°12 page 2.41 un exemple de disposition sur un bordereau.

II. 5 - EXEMPLE DE DESCRIPTION D'UN DOCUMENT.

Dans les pages suivantes le lecteur trouvera d'abord le texte d'un article dans son intégralité, puis sa description à l'aide de notre langage.

Hemodynamics and coronary blood flow with counterpulsation

JOHN A. DORNANDY, M.D.
ROBERT H. GOETZ, M.D.*
DAVID KRIPKE, M.D.

BRONX, N. Y.
From the Department of Surgery,
Albert Einstein College of Medicine

Post-surgical counterpulsation which consists of alternate withdrawal and rejections of a specified volume of blood synchronously with but counter to the patient's cardiac action is perhaps the most attractive system of mechanical assistance so far devised. By withdrawing blood during systole and returning it during diastole at increased pressure, counterpulsation should achieve the goal of most types of mechanical assistance to the circulation, i.e., reduce the work load of the left ventricle and improve coronary circulation. Data available, however, fail to present a clear picture of either the physiological effect of counterpulsation or its therapeutic potential. Increases in coronary flow up to 30 percent were demonstrated by some^{1,2} but could not be confirmed by others.^{3,4} Similar discrepancies have been reported in the response of other hemodynamic parameters. Little attention has been paid in the literature to the fact that pressure and flow characteristics of the coronary tree differ in normo- and hypotensive states. Since many of the patients in whom counterpulsation may be indicated presumably will be hypotensive, it appeared important to investigate the effect of counterpulsation, not only in the normo- but also in the hypotensive state and to compare the re-

sponse in coronary blood flow and left ventricular work in the 2 groups. In order to overcome the difficulties encountered with high irregular heart rates, counterpulsation was combined in some experiments with R wave triggered pacing. The preliminary results were reported in the *Surgical Forum* 1967.⁵

METHODS AND MATERIALS

The pumping system. The mechanical counterpulsation system consists of 3 modules: 1, the electronic synchronizing unit; 2, the pneumatic control unit; and 3, the pumping chamber (Fig. 1). The electronic synchronizing unit controls duration and proper synchronization of the pump with the heart rate. An amplifier circuit picks up the triggering signal from either the R wave of the electrocardiogram or the systolic peak of the arterial pressure tracing, which is then passed on to the pneumatic control unit where it activates the input and output valves of the pump. The pump itself consists of a metal bellows in a pump cylinder. (The stroke volume can be adjusted from 0 to 75 ml. It is connected to one or both femoral arteries of the patient by thin-walled, stiff Teflon catheters which are advanced into the iliac arteries preferably just distal to the bifurcation of the aorta.)

Experimental design. Counterpulsation was applied in 56 experiments to 31 dogs weighing between 40 and 50 pounds. Anesthesia was induced with 3.8 percent thopental sodium and maintained by the infusion

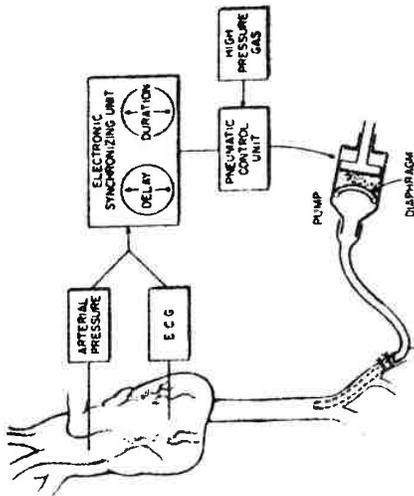


Fig. 1. Diagram illustrating principle of counterpulsation. Activated by the arterial pressure or R wave of the ECG, the essential components are the electronic synchronizing unit, the pneumatic control unit, and the pump.

of 2 percent Nembutal. In experiments requiring a thoracotomy an endotracheal tube was passed and connected to a positive pressure respirator using 10 percent oxygen. All animals were anticoagulated with 1.5 mg. heparin per kilogram.

Pressures were recorded by means of Stamm strain gauges through catheters placed in the aortic arch, right atrium, and right and left ventricles. Cardiac output was measured by dilution techniques after injection of 0.5 ml. of 5 percent Indocyanine green into the superior vena cava. The maximum rate of left ventricular pressure ascent during isometric contraction (maximum dp/dt) was measured directly from the left ventricular pressure tracings and expressed as millimeters Hg per second. Left ventricular work in gram-meters was calculated as the product of cardiac stroke volume and the planimetrically integrated mean systolic aortic pressure in millimeters Hg obtained from 50 mm. per second pressure tracings. Tension-time index in millimeters Hg per minute was

obtained as a product of mean systolic aortic pressure, duration of systole, and heart rate.

Blood flow was recorded in the carotid and anterior descending coronary arteries by means of an electromagnetic flowmeter. The flows were calculated by measuring the area under the continuous flow rate curve and applying a calibration constant. A flanged catheter was tied into the coronary sinus and the blood drained via a measuring cylinder into a reservoir and thence returned to the right femoral vein by a Dale Schuster pump. Coronary sinus flow was measured either directly by 30 second collections in the measuring cylinder or continuously by a pressure transducer connected to the base of the cylinder as described in the previous publication.⁶ Two sets of electrocardiogram leads were attached, one for activating the pump and the other for recording.

For the experiments, 25 animals were made hypotensive either by graded hemorrhage (Group 1) or left ventricular failure (Group 2). In the latter group a double

Supported by National Institutes of Health Grant No. NIH R 1343.

Submitted for publication Aug. 21, 1967; resubmitted April 5, 1968.

*Recipient of Fellowship of the Health Research Council of the City of New York under Contract No. 1-27.

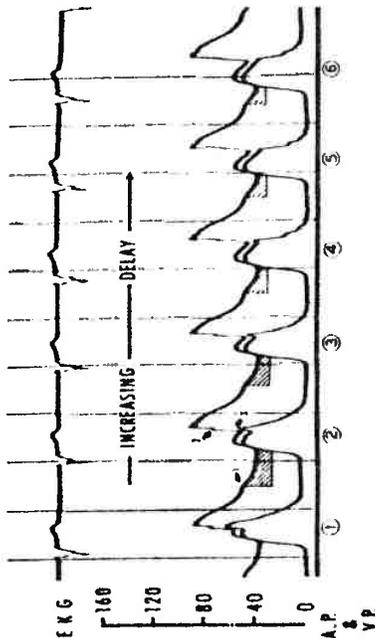


Fig. 2. ECG, central aortic pressure (A.P.), and left ventricular pressure (L.V.P.) simultaneously recorded during counterpulsation. The delay in the electronic synchronizing unit is gradually increased during the course of the tracing to illustrate the principle of proper phasing. Arrow 1 and shaded area marks the drop in aortic pressure during diastole when pump withdrawal begins. It is progressively shortened from pulse 1 to 6. Arrow 2 marks the rebound in the aortic tracing coincident with the closing of the aortic valves while Arrow 3 points to the pressure spike produced in the right ventricle by the premature reinjection. At pulse 4 reinjection occurs during ventricular aortic A.P. pulse 5 reinjection begins at closure of aortic valves.

lumen balloon catheter was passed retrograde through the right carotid artery into the left ventricle. One lumen was open at the tip of the catheter and was used for pressure recordings; the other was connected to a balloon which was placed in the left ventricle and inflated with 10 to 15 ml of saline to obstruct the aortic valve ring. As soon as the left ventricle showed signs of failure, usually after 20 to 60 seconds, the balloon was deflated and withdrawn. This left a damaged and overstretched left ventricular myocardium as judged by the persistent decrease in systemic blood pressure.

The hypotensive animals were then treated with counterpulsation and after all data had been obtained, the blood pressure was restored by reinflation of the shed blood or a constant drip of epinephrine. The effect of counterpulsation was then compared before and after normalization of blood pressure.

In 6 additional animals counterpulsation

ing of the pump. The beat-by-beat changes in end diastolic pressure with changes in the delay are well illustrated.

Reinjection should begin as early in diastole as possible because the necessary assistance is least in the immediate postsystolic phase. Premature injection while the aortic valves are still open hinders rather than facilitates ventricular emptying and produces a sharp rise in intraventricular pressure, best seen in the first beat of Fig. 2.

RESULTS

The results were divided into 3 groups: normotensive animals with aortic systolic pressures of over 100 mm Hg Group 1; hypertensive animals with aortic blood pressures below 100 mm Hg Group 2; and

animals in which counterpulsation was combined with R wave coupled pacemaker (Group 3). This last group contained both normotensive and hypertensive animals and will be considered separately. There were 25 experiments each in Groups 1 and 2 and 6 experiments in Group 3.

Blood pressures. The effects of counterpulsation on the blood pressure are well illustrated in Fig. 3, recorded from an animal which had been bled to hypotensive blood pressure levels. Aortic and left ventricular pressures were simultaneously recorded. The marked decrease in the systolic ejection pressure and the imposing height of the counterpulse in each diastole are well documented. These changes in aortic and left ventricular pressure pattern were the

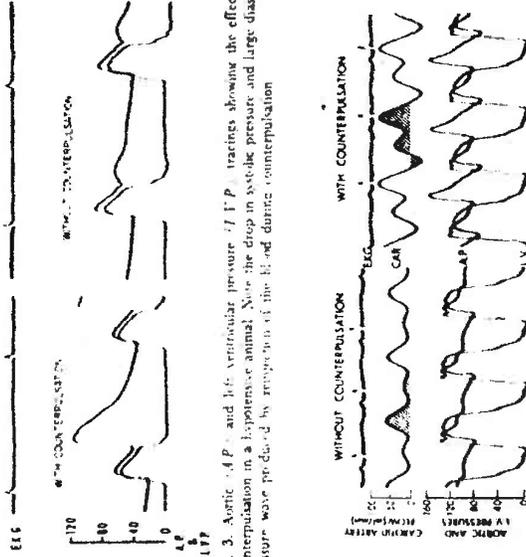


Fig. 3. Aortic (A.P.) and left ventricular pressure (L.V.P.) tracings showing the effect of counterpulsation in a hypotensive animal. Note the drop in systolic pressure and large diastolic pressure wave produced by retrograde flow during counterpulsation.

Fig. 4. Effect of counterpulsation on pulsatile arterial flow in the carotid artery (CAR). The shaded area under the flow rate curve represents the volume of bio-d flow per beat and is approximately the same before and after. With counterpulsation a brief period of retrograde flow occurs corresponding with the beginning of pump withdrawal in late diastole. With counterpulsation the peak flow is associated with early and mid-diastole.

same in both the normo- and hypotensive groups (Fig. 4). Peak systolic pressure dropped in the normotensive group by 11.2 percent and in the hypotensive group by 12 percent. The fall in end diastolic pressure was 12.4 percent and 19 percent, respectively (Table I). As is apparent from Fig. 3 the most significant change with counterpulsation concerned the peak diastolic pressure which was elevated by an average of 69 percent in the normotensive and 88 percent in the hypotensive group. The peak diastolic pressure is, without exception, significantly raised by counterpulsation, whereas peak systolic pressure decreases in both the hypotensive and normotensive groups.

Left ventricular systolic pressure changes paralleled in all experiments the decrease seen in the peak systolic pressure in the aorta.

Heart rate. The heart rate varied between 118 and 165 beats per minute and was not altered by counterpulsation.

Cardiac output. Cardiac output was measured in 22 experiments in Groups 1 and 2

and was found to be essentially unaltered by counterpulsation. Counterpulsation introduced considerable variation in the results of cardiac output as measured by the dye dilution technique and, therefore, small differences recorded were not thought to be significant.

Carotid blood flow. Carotid artery flow rates did not vary significantly with counterpulsation, although the phasing of the flow did change considerably, as can be seen in Fig. 4. The amplitude of the rate curve increased with counterpulsation but there was a period of actual retrograde flow late in diastole when the pump began to withdraw blood. The peak flow which normally coincides with the peak arterial pressure was shifted into diastole and became synchronous with the pump retraction phase. Thus, counterpulsation altered the pattern of carotid flow as well as carotid pressure.

Coronary sinus flow. Whereas the effect of counterpulsation on aortic and ventricu-

Table I

	Peak systolic aortic pressure (mm. Hg.)			Peak diastolic aortic pressure (mm. Hg.)		
	Control	Counter-pulse	Difference (%)	Control	Counter-pulse	Difference (%)
Normotensive	131.6	115.2	-11.2	95.4	159.0	+69.0 (72.3%)
Hypotensive	69.0	4.8	-12.0	47.6	83.4	+88.0 (119.0%)

Table II

	Left ventricular stroke work (Erg.-meters)			Maximum dp/dt (mm. Hg./second)		
	Control	Counter-pulse	Difference (%)	Control	Counter-pulse	Difference (%)
Normotensive	29.3	34.0	+16.2	3,270	3,870	+18.3 (2,828)
Hypotensive	8.27	6.26	-24.3	1,275	893	-31.0 (4,690)

lar pressures is of the same order in the normo- and hypotensive groups, marked differences were recorded in coronary sinus flow. The changes recorded in 38 experiments are plotted in Fig. 5. Average coronary flow increased by 24 percent in the normotensive group, but it remained essentially unchanged in hypotensive animals (Table I).

Plasma hemoglobin and pH. In 4 experiments the plasma hemoglobin and pH were measured repeatedly during counterpulsation. In the first 2 experiments, the synchronization and the phasing of the pump were imperfect, and free hemoglobin rose to 80 and 90 mg. percent, respectively, after 2 hours. In the other 2 experiments the timing of the pump was optimal and the plasma hemoglobin levels reached 36 and 40 mg. percent after 2 hours of counterpulsation. A mild acidosis was observed in all experiments although this is unlikely to be related to counterpulsation.

Left ventricular function. Of the various parameters used for assessing left ventricular function only left ventricular stroke work differed significantly in response to counter-

pulsation in both the normo- and hypotensive groups. Left ventricular stroke work decreased 16 percent in the normotensive and 24 percent in the hypotensive group which was significant at the 90 percent level (Table II). The time-tension index fell with counterpulsation in all experiments by an average of 16 percent, and maximum rate of left ventricular pressure ascent fell by 29 percent, with no significant difference between the hypotensive and the normotensive groups.

R wave coupled pacing. At the high heart rates usually seen in dogs, difficulties may be encountered in withdrawing sufficient blood during systole and placing the counter-pulse exactly in early diastole. This can be overcome by combining R wave coupled pacing with counterpulsation, as illustrated in Fig. 6. R wave coupled pacing halved the heart rate from 160 per minute to 80 per minute and increased the length of diastole, permitting exact placing of the counter-pulse early in diastole. Combining counterpulsation with R wave coupled pacing did not significantly affect the results of counterpulsation alone.

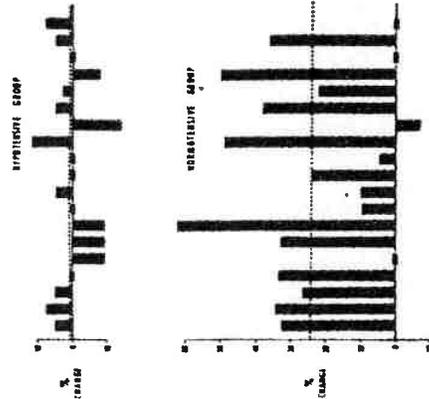


Fig. 5. Summary of changes in coronary flow with counterpulsation expressed as a percentage change of the control values. The dotted line indicates the mean change in the 2 groups and shows that while in the normotensive group average coronary flow increased by 24 percent, in the hypotensive group there was only an insignificant change.

End diastolic aortic pressure (mm. Hg)				Mean aortic pressure (mm. Hg)				Coronary sinus flow (ml. minute)			
Control	Counter-pulse	Difference (%)		Control	Counter-pulse	Difference (%)		Control	Counter-pulse	Difference (%)	
95.4	83.4	-12.4 (12.2%)		116.5	120.7	+3.0		89.2	111.6	+24 (4.7%)	
47.5	38.6	-18.0 (37.1%)		54.6	46.3	-14.0		49.7	35.4	-11.0	

Transmension index (mm. Hg minute)				Peripheral:arterial resistance (dyne second/cm ²)			
Control	Counter-pulse	Difference (%)		Control	Counter-pulse	Difference (%)	
2,686	2,296	-14 (2.801)		3,076	3,114	+3	
781	644	-18 (2.449)		1,950	1,662	-15 (2.2%)	

DISCUSSION

Left ventricular function and coronary blood flow. One of the basic aims of counterpulsation is to decrease left ventricular work, by reducing the left ventricle to a chamber which merely conveys blood from the pulmonary system to a systemic pump, i.e., the counterpulsator. In our experiments there was a consistent fall in all 3 parameters of left ventricular function: stroke work, time-tension index, and maximum dp/dt. Counterpulsation, therefore, decreases the work load of the left ventricle by any criteria. Moreover, this decrease is achieved without loss of peripheral perfusion as evidenced by the unaltered cardiac output and carotid artery flow. The decrease in the work of the left ventricle was seen both in the normotensive and hypertensive preparations.

To justify counterpulsation as an effective form of assistance to the failing circulation it is essential to establish that it increases coronary flow. Our results in this area were the most striking. Coronary blood flow was significantly increased by counterpulsation if the central aortic pressure was normal.

but remained unchanged in the presence of hypotension. The dependence of the change in coronary flow on central aortic tension is particularly interesting in view of the fact that peak diastolic pressures were raised to equal degrees by counterpulsation in both the hypotensive and normotensive groups. Thus recording adequate pressure increments in diastolic pressure during counterpulsation does not necessarily imply that they are accompanied by corresponding increases in coronary flow. Our data are quite consistent and offer an explanation for the wide discrepancies reported in the literature. Sroff and associates¹⁰ and Goldfarb¹¹ reported a 50 percent and 31 percent increase in coronary flow, respectively, but this was not confirmed by Wilman and associates¹² and LaFemine.¹³ Jacoby¹⁴ describes an increase in coronary anatomoses following counterpulsation in dogs subjected to coronary embolization with plastic spheres through he did not show prepump arteriograms.

The relation between coronary flow and central aortic tension appears to be a reflection of the "windkessel" function of the

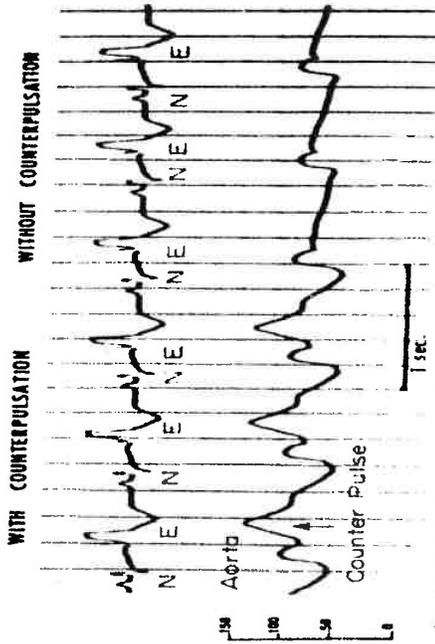


FIG. 6. R wave coupled pacing with counterpulsation. mechanical heart rate is half spontaneous rate. Note the aortic diastolic pulse wave. F, counterpulsation and considerable drop of end diastolic pressure. A, arterial aortic beat. E, mechanically ineffective extrasystole.

applies pneumatic power linkage. In contrast the system developed by Callaghan, Chesnut, and Watkins¹⁵ uses hydraulic coupling which has a theoretical advantage which is due to the incompressibility of liquid. However, we found no differences in the efficiencies of both systems even with heart rates of 160 per minute. Lastly, proper cannulation of the arterial system becomes important in order to avoid sudden changes in the lumen size and so to prevent hemolysis from cavitation and turbulence. Teflon tubing was found to be most suitable because it could be molded and because it allowed the lowest ratio of external to internal diameter while maintaining its comparative inelasticity.

To achieve maximum effectiveness the pump stroke should be delivered through a large cannula as proximally in the aorta as possible. Under clinical circumstances a peripheral artery has to be used for cannulation and the size of that artery limits the size of the cannula. The subclavian

artery was used experimentally and found unsuitable. Many investigators used direct cannulation of the aorta^{1,2,3} which necessitates a laparotomy and detracts from the clinical applicability. We cannulated the superficial femoral artery. This may present considerable difficulties in the elderly arteriosclerotic patient. Therefore methods of applying counterpulsation without cannulating a vessel by encasing the legs in an airtight pressure cuved cylinder have been devised.^{4,5} Birnwell and his group⁶ also developed a method of counterpulsation to assist the right side of the heart, using a cuffed endotracheal tube to apply a counterpulsating pressure through the trachea.

Combined R wave coupled pacemaking and counterpulsation. Although we found it possible to use counterpulsation effectively, even with heart rates as high as 160 per minute, pacing and synchronization may present difficulties with the high heart rates or disturbances in rhythm found clinically in shock. Shook⁷ and Willman and associates⁸ solved the problem in the experimental preparation by inducing heart block and pacing the dogs at a slower controlled rate, while Ellis⁹ limited pump response to alternate cardiac cycles only. We employed R wave coupled and paired pacing to overcome this difficulty. Gaining the electronic synchronizing mechanism in such a way that it was not triggered by the extrasystole proved to be no problem and demonstrated that combination of both modalities is possible. However, our results with combined use of counterpulsation and synchronous pacing did not demonstrate any improvement over those with counterpulsation alone.

Clinical application. Theoretically counterpulsation is most suited to the treatment of shock caused by myocardial infarction or anoxia, i.e., circulatory failure caused by left ventricular disease. Approximately 8 to 10 percent of patients develop protracted hypotension following acute myocardial infarction,^{10,11} and it is in these cases that a peripheral circulatory support, which at the same time decreases left ventricular work and increases coronary perfusion, may be

useful. In view of the evidence that secondary heart failure plays an important role in making some forms of hemorrhagic or septic shock irreversible, it is possible that counterpulsation may also have a part to play in the treatment of this condition. In addition there is some evidence that counterpulsation allows the rapid transfusion of blood without a significant rise in right atrial pressure in patients with borderline compensation of congestive heart failure. Finally, counterpulsation may be a convenient way to control failure following cardiac surgery.

SUMMARY

Counterpulsation, or postastolic augmentation, is a form of mechanical assistance to the failing circulation depending on the withdrawal of blood from the central arterial system during cardiac systole and its reinjection during diastole. It lowers systolic ventricular ejection pressure and at the same time maintains peripheral flow and increases the diastolic pressure. Counterpulsation decreased all the parameters of left ventricular work, i.e., the tension-time index, stroke work, and maximum dip dt. Peripheral perfusion as gauged by carotid flow and cardiac output was maintained. However, coronary flow increased only in dogs with a normal central aortic pressure. Despite similar increases in peak diastolic pressures counterpulsation did not produce a rise in coronary flow in hypotensive animals. When combined with R wave coupled pacemaking in order to facilitate synchronization, counterpulsation did not produce significantly different results. The clinical implications of these findings are discussed.

REFERENCES

- Birnwell, W. C., Sorloff, H. S., Sachs, B. S., Levine, H. J., and Detelting, R. A., Jr.: Assisted circulation. V. The use of the lungs as a pump. A method for assisting pulmonary blood flow by varying airway pressure synchronously with the ECG. *Tr. Am. Soc. Artif. Int. Organs* 9:192, 1963.
- Callaghan, P. B., Chensut, M. G., and Watkins, D. H.: A computer-automated mechanism for assisted circulation. *Tr. Am. Soc. Artif. Int. Organs* 11: 36, 1965.
- Ellis, P. R., Lee, C., Wong, S. H., Del Rosario, V. C., Hlavnd, J. W., and Prater, G.: Assisted circulation in the treatment of experimental heart failure. *Arch. Surg.* 90: 879, 1965.
- Goetz, R. H., Selmonsky, C. A., and Staele, D.: The effect of the amine buffer tin amino methane on the renal blood flow during hemorrhagic shock. *Surg. Gynec. & Obst.* 117: 715, 1963.
- Goldfarb, D., Kniecherbocker, G. G., Jude, J. R., and Bahnon, H. T.: Treatment of circulatory failure with a new system for selective diastolic augmentation. *Stavosky* 56: 547, 1964.
- Gouldstein, J. V., Dormandy, J. A., Kriplek, D. C., and Goetz, R. H.: Coronary sinus blood flow in counterpulsation. *Surg. Forum* 18: 137, 1967.
- Hever, H. E.: A clinical study of shock occurring during acute myocardial infarction. An analysis of 58 cases. *Am. Heart J.* 62: 416, 1961.
- Jacobey, J. A., Taylor, W. J., Smith, G. T., Gorlin, R., and Harken, D. E.: A new therapeutic approach to acute coronary occlusion. II. Opening dominant coronary collateral channels by counterpulsation. *Am. J. Cardiol.* 11: 218, 1963.
- Lafontaine, A. A., Law, H. B., Cohen, M. C., and Harken, D. E.: Assisted circulation. II. The effect of heart rate on synchronized arterial counterpulsation. *Am. Heart J.* 64: 779, 1962.
- Malarb, M., and Rosenbergs, B. A.: Acute myocardial infarction in a city hospital. III. Experience with shock. *Am. J. Cardiol.* 35: 487, 1960.
- Osborne, J. J., Russi, M., and Satal, A.: Circulatory assistance by external pulsed pressures. *Am. J. M. Electron.*, April-June, 87, 1964.
- Shenk, W. G., Dolon, N. A., Camp, F. A., McDonald, K. E., Pollack, L., Gage, A. A., and Chardak, W. M.: An experimental evaluation of counterpulsation and left ventricular bypass. *Arch. Surg.* 88: 327, 1964.
- Soroff, H. S., Levine, H. J., Sachs, B. F., Birnwell, W. C., and Detelting, R. A., Jr.: Assisted circulation. II. Effects of counterpulsation on left ventricular oxygen consumption and hemodynamics. *Circulation* 27: 222, 1963.
- Watkins, D. H., and Duchesne, E. R.: Post-systolic myocardial augmentation: Physiological considerations and techniques. *J. Appl. Physiol.* 17: 61, 1962.
- William, V. L., Cooper, T., Robert, A., and Harken, C. R.: Cardiac assistance by diastolic augmentation. Hemodynamic evaluation on dogs with complete heart block. *Tr. Am. Soc. Artif. Int. Organs* 7: 198, 1961.

```

00 $1002*002*1969*065*0311*0320* HEMODYNAMICS AND CORONARY BLOOD FLOW WITH
01 COUNTERPULSATION/
02 AUT=MAIN=DORMANDY(JA)#COWR=GOETZ(RH),KRIPKE(O)/
03 MEASU=PRESS=AO,RV,RA,LV #
04 BFLOW=HEART(DYE),CORON(SINUS),CAROTID,CORON(ART,ANTERIOR)/
05 DDCU =FIG=2 ; PAP=ENG ; TABLE=2 ; REF=14 /
06 PATT =SPEC=D0G ; NBR=25 ; WGT=<18:22.6> ; EXP=1#
07 MODIF=1 ; NBR=6 ; EXP=2 /
08 ANEST=KIND=GENERAL ; THIOP=I ; NEMB=I /
09 COUNP=STVDL=<0:75> ; CATCO=TEFLON ; CATEN=AORTA(BIFURCATION) ; TRIG=PRESSURE /
10 ACBAG=HEPAR=1.5 ; TIME=<1 /
11 VENT =MODE=MECHANICAL ; TYPE=PP /
12 CHRON=EVENT=OP , COUNP , I , COUNP ;
13 REP =1 , <2:3> , <3:4> , <4:5> /
14 PROC= COND=HYPOTENSION(HEMORRHAGE) ; AOSP=<100 ; TIME=<1:3> #
15 MODIF=1 ; COND=HYPOTENSION(FAILURE(VENTRICULAR)) #
16 BLOOD=I(REINFUSION) ; TIME=<3:4> #
17 EPIN=I(PERFUSION) ; TIME=>3 #
18 COND=NORMOTENSION ; AOSP=>100 ; TIME=>4 #
19 EXP=2 ; HRATE=80(PACING) /
20 RESUL=TIME = <1 , <2:3> , <3:4> , <4:5> ;
21 AOSP = 69.0 , 60.8 , 141.6 , 125.2 ;
22 AOP = 47.6 , 83.4 , 95.4 , 159.0 ;
23 AODP = 47.6 , 38.6 , 95.4 , 83.4 ;
24 MAP = 54.6 , 46.3 , 116.5 , 120.7 ;
25 CSFLO = 49.7 , 50.4 , 89.2 , 111.6 ;
26 LVSTW = 8.23 , 6.20 , 29.3 , 24.6 ;
27 MDPDT = 1255 , 893 , 520J , 3870 ;
28 TTI = 781 , 644 , 2686 , 2296 ;
29 PVRES = 1950 , 1662 , 3006 , 3114 #
30 EXP=1 ; HRATE=<118:185> #
31 *FOR EXP=2 COMBINING COUNP WITH R WAVE COUPLED PACING DID NOT SIGNIFICANTLY
32 AFFECT THE RESULTS OF COUNP ALONE , CO UNALTERED BY EXP=1#

```

CHAPITRE III

CHAINE D'ACQUISITION DES DOCUMENTS

III. 1 - INTRODUCTION

Le chapitre précédent nous a appris à décrire les documents à l'aide d'un langage de description et à les présenter sur des bordereaux. Cette forme des documents est dite externe.

La chaîne d'acquisition des documents a pour but de ranger ces documents (après perforation sur cartes - fichier FCARTE) dans un fichier FDOCD (fichier définitif des documents) sous une forme que nous appellerons interne.

Le passage de la forme externe à la forme interne d'un document est effectué par un programme de traduction (de nom TRD). Ce programme détecte les erreurs commises par l'utilisateur et ne conserve que les documents non erronés dans un fichier de transition FDOC. Un tel document peut encore contenir des erreurs que le programme ne peut détecter (par exemple 3.01 au lieu de 30.1), Celles-ci sont corrigées sur la forme externe (cartes) qui est ensuite soumise à une nouvelle traduction. La nouvelle forme interne ainsi obtenue va automatiquement remplacer l'ancienne version sur le fichier FDOC.

Les documents de FDOC sont périodiquement transférés sur le fichier FDOCD (fichier définitif) par un programme de nom TRF.

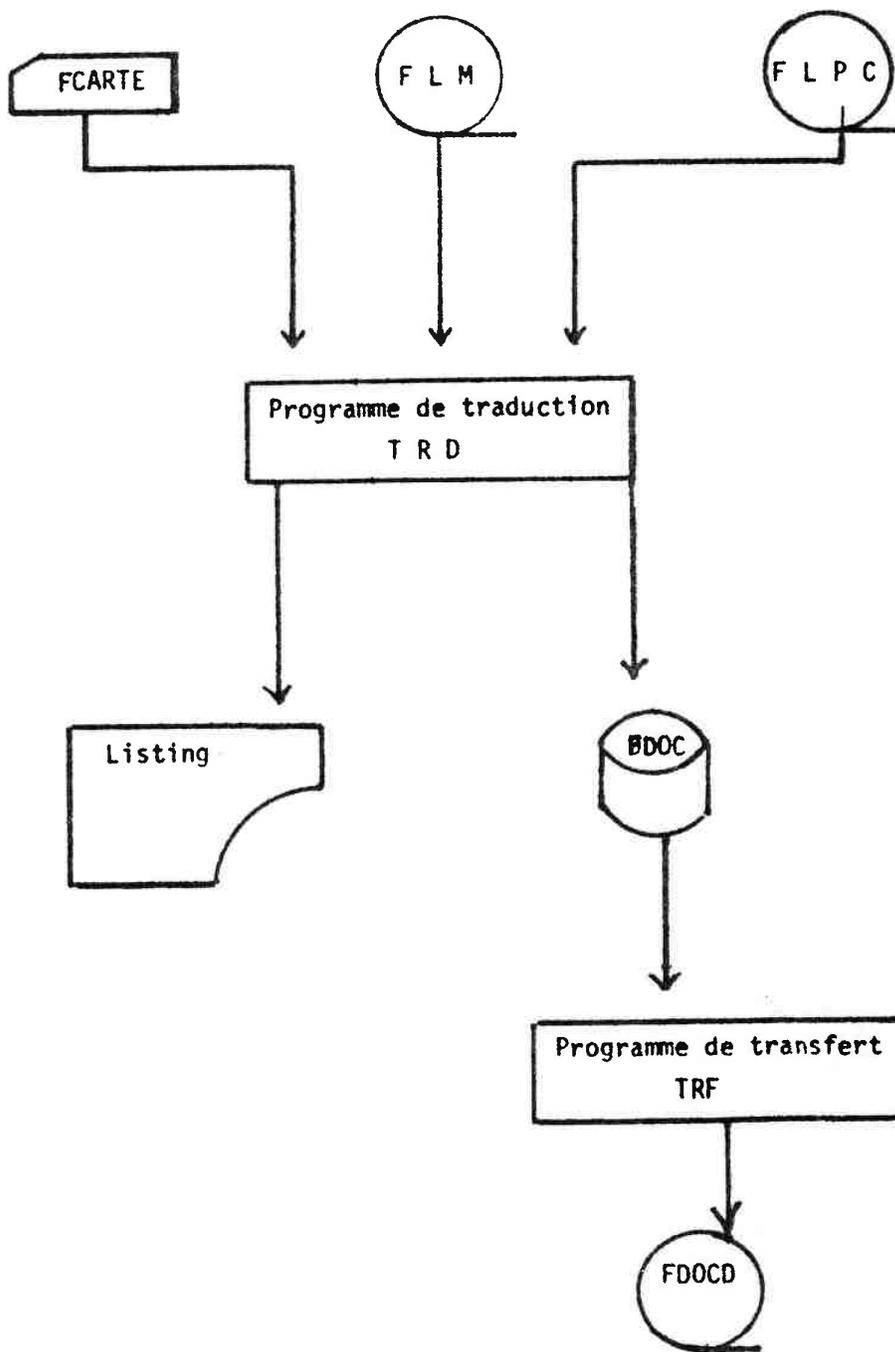
La description détaillée de la forme interne des documents et des fichiers FDOC et FDOCD est donnée en III.3 - C.

Le programme de traduction fait des consultations du lexique des noms de paragraphes et de composantes (fichier FLPC) et du lexique des mots (FLM). Ces deux lexiques seront décrits dans la troisième partie de ce chapitre consacrée aux fichiers. En particulier nous parlerons de la création et la mise à jour de ces fichiers.

Les directives d'utilisation de cette chaîne d'acquisition des documents sont exposées en III. 2 et intéressent tout particulièrement les utilisateurs.

Le schéma n° 1 de la page 3.2 montre l'organisation générale de la chaîne.

Schéma n° 1 : Organigramme de la chaîne.



III. 2 - DIRECTIVES D'UTILISATION DE LA CHAÎNE.

A - Les tâches à entreprendre par l'utilisateur.

Nous nous adressons ici à l'utilisateur qui, ayant décrit sur des bordereaux un ou plusieurs documents, veut les faire entrer en machine.

Après perforation les documents sont mis dans le fichier d'entrée FCARTE dont nous avons parlé dans l'introduction.

Le programme de traduction (TRD) les prend en charge et produit un listing à l'usage de l'utilisateur. Celui-ci y trouvera l'impression de la forme externe des documents (celle qui se trouve sur cartes) ainsi que des messages que nous explicitons plus loin (III. 2 - B).

Le dépouillement de ce listing conduit à deux tâches :

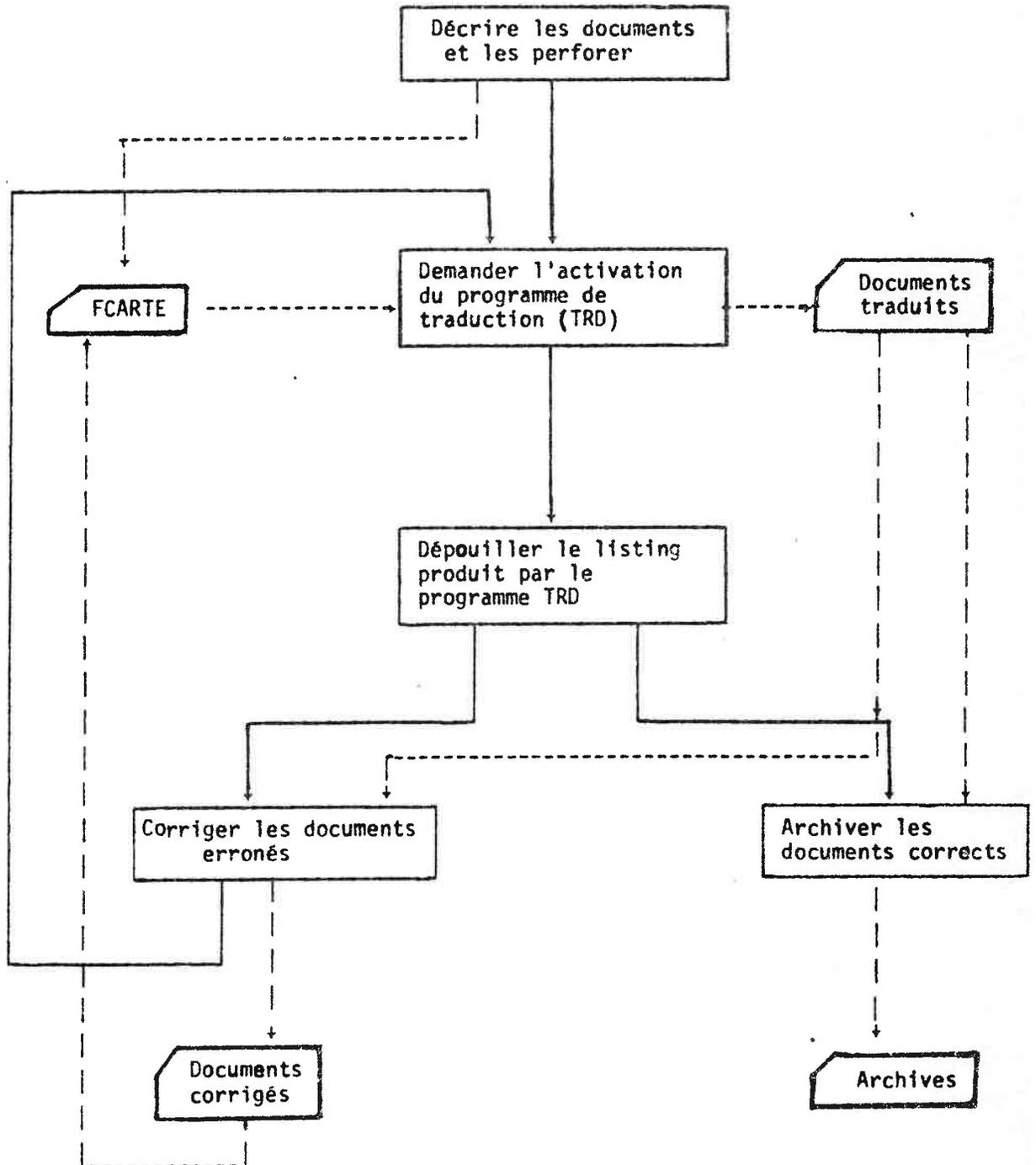
- corriger les cartes qui correspondent à des documents erronés (que ce soit des erreurs de syntaxe signalées par le programme ou des erreurs sémantiques détectées par l'utilisateur). Ces documents sont remis dans le fichier d'entrée FCARTE pour une nouvelle traduction.

- archiver les cartes qui correspondent aux documents corrects (c'est-à-dire les mettre dans un bac réservé à cet usage).

Le schéma n° 2, page 3.4, est destiné à l'utilisateur et représente les tâches qu'il a à entreprendre. Ce schéma n'est pas un organigramme classique et ne fait pas intervenir les fichiers dont nous avons parlé dans l'introduction et dont l'usage est transparent à l'utilisateur.

Les flèches pleines matérialisent la marche à suivre. Les flèches en pointillé montrent le trajet des documents (sur cartes) entre différents bacs (représentés par le dessin d'une carte). Les tâches sont encadrées.

Schéma n° 2.



B - Description des tâches.

Nous décrivons ici les tâches représentées sur le schéma n° 2 de la page 3.4

B - 1. Décrire les documents et les perforer.

Tout le chapitre II est consacré à ce travail de description.

B - 2. Demander l'activation du programme de traduction.

Cette demande est à faire auprès de l'"informaticien".

B - 3. Dépouiller le listing produit par le programme TRD.

Le texte des documents (celui qui est sur cartes) est restitué sur ce listing. A ce texte s'ajoutent des messages émis par le programme et que nous allons étudier maintenant.

B - 3. 1 - Messages d'erreur de syntaxe.

Ces messages suivent la ligne dans laquelle les erreurs ont été détectées. Pour faciliter le travail de dépouillement, ces messages ne sont pas codés et indiquent l'endroit où l'erreur a été trouvée.

Nous donnons en exemple un petit extrait de listing :

```
04 PATT = SPEC = DØG (MØNGREL ; WGHT = 1Ø ; EXP = 1 #
ERREUR DE PARENTHESSES.....DØG (MØNGREL
ERREUR DANS PARTIE ENTIERE....1Ø

05 SPEC = DØG (BEAGLE) ; WGHT = >10:15> /
ERREUR DANS UNE FOURCHETTE..... >10:15>
```

- Le premier message signale qu'il manque une parenthèse fermante après le mot MONGREL.

- Le deuxième message signale une erreur dans le nombre 1Ø. Le programme y a en effet trouvé la lettre "Ø". Il y a eu une erreur de perforation (lettre "Ø" au lieu du chiffre zéro). Pour éviter ce genre d'erreurs, nous demandons aux utilisateurs de barrer le caractère "Ø" pour désigner la lettre afin de la distinguer du zéro (quand il écrit sur des bordereaux).

- Le troisième message signale une fourchette incorrecte. Le premier symbole doit en effet être < et non pas >.

Nous ne pensons pas qu'il soit utile d'expliciter ici tous les messages de ce type. Leur texte est assez précis pour que l'utilisateur puisse remédier sans trop de peine à ses erreurs (en se reportant éventuellement aux cartes syntaxiques du chapitre II).

Signalons que si plusieurs erreurs sont découvertes dans une même ligne, elles font toutes l'objet d'un message (voir l'exemple).

Lorsque l'utilisateur a corrigé ses erreurs, il demande une nouvelle traduction de son document (voir le schéma n° 2, page 3.4).

B - 3. 2 - Message : "MOT INCONNU".

Nous rappelons que l'utilisateur est tenu d'employer les mots consignés dans le lexique dont une copie écrite est à sa disposition.

Le programme de traduction signale les mots du document qui ne figurent pas dans le lexique. Le message édité suit la ligne où le mot inconnu a été trouvé.

Exemple.

```
08 PROC = AOSP = <100 ; COND = HEMORHAGE #
MOT INCONNU .....HEMORHAGE
Ce mot s'écrit en effet avec deux "R".
```

Un tel message recouvre en fait trois cas :

1. Le mot signalé est mal orthographié par l'utilisateur ou a été mal perforé (erreur de frappe). L'utilisateur corrige la carte correspondante.
2. Le mot signalé est bien celui que l'utilisateur a voulu employer mais il ne figure pas encore dans le lexique des mots. Il demande son introduction dans le lexique (il aurait dû le faire avant de l'utiliser, voir III. 3 - B).
3. Le mot signalé est synonyme d'un mot qui existe déjà. L'utilisateur est invité à utiliser le mot existant.

B - 3. 3 - Messages concernant les noms de paragraphes et de composant

Les erreurs commises dans l'écriture des noms de paragraphes ou de composantes sont signalées par un message suivi du nom erroné.

Exemple.

04 PAT = SPEC = DOG ; WGHTT = 10 ; AGE = 4 #

NOM DE PARAGRAPHE INEXISTANT PAT

NOM DE COMPOSANTE INEXISTANT..... WGHTT

Les noms corrects sont respectivement PATT et WGHT.

Une autre erreur consiste à utiliser un nom de composante dans un paragraphe où elle ne peut figurer.

Exemple.

06 VENT = MAIN = KANTROWITZ ; MODE = MECHANICAL /

COMPOSANTE INTERDITE DANS CE PARAGRAPHE MAIN

L'utilisateur a utilisé MAIN (nom d'auteur principal). Il voulait probablement donner le nom de l'auteur de l'appareil de ventilation qui se donne avec la composante MARK.

B - 3. 4 - Messages de fin de document.

Après la fin du texte de chaque document l'un des deux messages suivant est imprimé :

1. *** DOCUMENT INCORRECT *** INDICATIF : xxxx

Le programme a détecté une ou plusieurs erreurs dans le document dont l'indicatif est xxxx. La forme interne pour le document en question n'est pas conservée dans le fichier FDOC (ce qui exclut automatiquement la conservation par mégarde d'un document erroné). Les erreurs signalées doivent être corrigées et le document soumis à une nouvelle traduction.

2. *** DOCUMENT CORRECT *** INDICATIF : xxxx

Le listing d'un document est suivi de cette mention quand le programme de traduction n'y a trouvé aucune erreur. Un tel document est rangé (sous sa forme interne) dans le fichier des documents (FDOC).

L'utilisateur est invité à relire le document afin de détecter d'éventuelles erreurs de description ou de perforation en se reportant au texte intégral de l'article (exemple d'erreur : 10.5 au lieu de 1.05). En particulier, la partie identification doit être lue afin d'y vérifier l'exactitude des renseignements qui y sont donnés (référence, année,... etc).

Si le document est jugé correct après cet examen, l'utilisateur range les cartes correspondantes dans les archives. C'est le seul travail à entreprendre dans ce cas. Si par contre l'utilisateur veut modifier une information, il peut le faire et remettre le document rectifié dans le fichier de départ (FCARTE). La nouvelle traduction obtenue remplace automatiquement la version précédente dans le fichier des documents (FDOC) en signalant l'opération par un message (voir B - 3.5), sauf si en faisant ses corrections l'utilisateur a introduit une erreur (le remplacement n'a alors pas lieu).

B - 3. 5 - Message "DOCUMENT_REMPLACE".

Ce message ne peut apparaître qu'après le message "DOCUMENT CORRECT". Sa forme exacte est la suivante :

*** DOCUMENT_REMPLACE *** INDICATIF : xxxx

Ce message intervient dans deux cas :

1. L'utilisateur remplace un document déjà enregistré dans le fichier FDOC par une nouvelle version (voir P - 3. 4).
2. L'indicatif que l'utilisateur a attribué à un nouveau document existe déjà dans le fichier des documents. Dans ce cas, il aura écrasé un document et commet une erreur qui peut être réparée comme suit :
 - le document correspondant à l'indicatif signalé par le message est recherché dans les archives (cartes).
 - le document qui est cause de l'erreur est corrigé (attribution d'un indicatif correct).
 - les deux documents sont mis dans le fichier FCARTE pour une traduction (le document détruit est donc reconstitué).

B - 3. 6 - Messages concernant le fichier des documents sur cartes (FCARTE).

Rappelons que pour chaque document les cartes sont numérotées à partir de zéro. Si une erreur de séquence se produit, elle est signalée par le message :

ERREUR DE NUMEROTAGE

Le document en cours de traitement est abandonné et les cartes rejetées sont listées accompagnées de la mention :

RECHERCHE DEBUT ENREGISTREMENT.

La traduction des documents qui suivent est entreprise.

Un autre message signale quant à lui une erreur de fin de fichier, c'est-à-dire que le dernier document ne se termine pas correctement. Le libellé du message est :

ERREUR FIN DE FICHIER.

B - 3. 7 - Messages concernant le programme de traduction.

Ces messages n'intéressent pas directement l'utilisateur. Ils signalent des débordements de tableaux ou de piles auxquels l'informaticien peut apporter des remèdes et non l'utilisateur. Si un tel message est imprimé, le document est suivi par la mention "DOCUMENT INCORRECT" et doit être à nouveau traduit après une petite intervention (augmenter des bornes de tableaux) sur le programme TRD.

Remarque à propos du fichier cartes.

Le fichier FCARTE peut contenir un nombre quelconque de documents d'utilisateurs différents et aussi bien des documents qui vont subir leur première traduction que des documents qui ont déjà été traduits mais pour lesquels une erreur a été détectée (soit par la machine, soit par l'utilisateur). Nous permettons ainsi aux différents utilisateurs de travailler indépendamment les uns des autres dans la mesure où ils peuvent à tout moment faire entrer un document en machine.

III. 3 - LES FICHIERS.

Dans cette partie, nous nous intéressons aux différents fichiers déjà mentionnés.

L'utilisateur peut très bien se passer de la lecture des parties techniques décrites sous la rubrique "Description" pour chacun des fichiers. Il est par contre invité à lire les rubriques "Création" et "Mise-à-jour".

Les noms attribués aux différentes zones sont ceux utilisés dans les programmes. La taille et le type de ces zones sont donnés à l'aide des notations COBOL, langage dans lequel sont écrits les programmes.

X(n) : zone de n caractères alphanumériques

COMP : entier binaire sur un mot machine

COMP-1: nombre flottant sur un mot machine.

9 (n) : zone de n caractères numériques.

A - Le lexique des noms de paragraphes et de composantes (FLPC).

A - 1. Généralités.

Les noms de paragraphes et de composantes sont répertoriés dans ce lexique dont l'utilisateur dispose d'une copie (sur papier) à utiliser pour décrire un document.

Ce lexique doit permettre quatre opérations :

- a) Trouver le code d'un nom de paragraphe ou de composante (coder).
- b) A partir d'un code (nombre entier) trouver le nom de paragraphe ou de composante qui lui correspond (décoder).
- c) Pour chaque nom de composante, donner la liste des noms de paragraphes où il peut être utilisé.
- d) Pour chaque nom de paragraphe, donner la liste des noms de composantes qui y sont rattachés.

Ces deux derniers points permettent d'une part de faire des vérifications dans la phase de traduction des documents et d'autre part d'améliorer le programme de recherche (interrogation).

A - 2. Description du fichier FLPC

Ce fichier est à accès séquentiel et son support est la bande magnétique. Comme la consultation des noms de paragraphes et de composantes est fréquente, nous avons décidé que ce fichier serait intégralement en mémoire centrale lors de son utilisation. Pour le moment, il ne comporte qu'un seul enregistrement logique constitué des tableaux présentés ci-dessous. S'il s'avère que cet enregistrement unique devient trop grand, il sera "éclaté" sur plusieurs enregistrements (problème lié à la taille des blocs physiques autorisée sur bande).

Tableau TLP (Tableau Lexique des Paragraphes)

Chaque nom de paragraphe occupe une ligne de ce tableau. L'indice de la ligne constitue le code du paragraphe. D'où un décodage efficace.

ligne j :

RESUL

X(5)

Le code de RESUL est j.

Tableau TLPA.

Les noms de paragraphes sont rangés dans ce tableau par ordre alphabétique afin de réaliser efficacement un codage par recherche dichotomique.

Le dessin d'une ligne de TLPA est le suivant :

ligne j :

nom de paragraphe	code
-------------------	------

X(5)

COMP

A remarquer que le code n'est pas conservé sous forme de caractères afin d'améliorer le temps d'exécution des programmes.

Tableau TLC (Tableau Lexique des Composantes) et TLCA.

: Le mode de rangement des noms de composantes se fait de manière analogue :

ligne j de TLC :

SPEC

 le code de SPEC est j

ligne i de TLCA :

nom de composante	code
-------------------	------

X(5)

COMP

Tableau MAT (Matrice)

Ce tableau à deux dimensions a autant de colonnes qu'il existe de noms de paragraphes et autant de lignes qu'il existe de noms de composantes. La taille d'un élément du tableau est d'un caractère (X). Si le nom de composante de code i est autorisé dans le paragraphe de code j, l'élément (i, j) contient 1 sinon 0.

Ce tableau répond à la fois aux exigences c) et d) formulées ci-dessus.

Dans l'avenir, il faudra transformer cette matrice d'octets en matrice de bits pour gagner de la place (Notre point de vue a été d'écrire les programmes rapidement dans un langage approprié (COBOL) qui ne permet pas la manipulation au niveau du bit).

Dans ce fichier sont en outre conservés les nombres de noms de paragraphes et de composantes dans deux mots (sous forme d'entiers COMP)

RTLP : remplissage de TLP

RTLC : remplissage de TLC

Les tableaux sont surdimensionnés de façon à pouvoir entrer 50 noms de paragraphes et 200 noms de composantes. Actuellement nous avons une vingtaine de noms de paragraphes et environ 150 noms de composantes (on pourra bien sûr augmenter les limites prévues si nécessaire).

A - 3. Création du fichier FLPC.

Les noms de paragraphes et de composantes choisis par les utilisateurs sont écrits sur les bordereaux de perforation conformément au dessin qui suit. Les lettres "b" représentent des blancs. Les noms de paragraphes et de composantes doivent être cadrés à gauche (colonne 5).

Le numérotage correspond en fait aux codes des noms de paragraphes et de composantes. Ces numéros sont surtout utiles en cas de destruction accidentelle du fichier FLPC qui peut être reconstruit très facilement et sans aucun risque d'erreurs (attribution d'un autre code à un nom par interversion de deux cartes ce qui serait catastrophique pour les documents déjà codés).

1	2	3	4	5					11		21		29		80
0	0	0	b	P	A	R	A	b	b						
0	0	1		noms de pa-		ragraphes									
numé-			rotage												
0	0	0	b	C	O	M	P	b	b						
0	0	1		noms de		composantes									
0	0	8		N	A	b	b	b		*			*		

Pour exprimer qu'une composante est autorisée dans le paragraphe de code k , il suffit de mettre une étoile dans la colonne de rang $k + 10$. Dans le dessin ci-dessus, la composante NA peut être utilisée dans les paragraphes de code 11 et 19. Notons que cette méthode qui peut sembler naïve, a donné satisfaction à l'utilisateur de par sa simplicité et sa clarté.

Le programme de création de ce fichier (son nom est LPC) ne présente aucun intérêt particulier. Il produit pour l'utilisateur un listing du lexique qui est présenté par noms de paragraphes. C'est-à-dire que pour chaque paragraphe sont donnés les noms de composantes qui y sont autorisés. Cette présentation simplifie bien sûr la tâche de l'utilisateur quand il décrit des documents.

B - Le lexique des mots.

B - 1. Généralités.

La constitution de ce lexique a été délicate et nous pensons qu'il est intéressant de montrer comment nous en sommes arrivés à la solution retenue.

A l'origine, nous avons pris pour option de construire ce fichier de façon automatique, c'est-à-dire que chaque fois qu'un mot nouveau (mot qui n'existe pas dans le lexique) apparaissait dans les documents, le programme de traduction TRD lui attribuait un code et le rangeait dans le lexique. Cette méthode a présenté plusieurs inconvénients :

- introduction dans le lexique de mots synonymes
- introduction dans le lexique du même mot sous des orthographes différentes (erreur de l'utilisateur ou de perforation)
- les deux points précédents sont encore bien plus graves si plusieurs utilisateurs décrivent des documents (chacun d'eux introduit alors "ses" mots).

Il était alors nécessaire d'avoir un programme de gestion de ce lexique des mots qui, entre autres opérations, permette de supprimer des mots. Mais la chaîne d'acquisition des documents en était considérablement alourdie pour éviter que les documents soient conservés en machine avec des codes correspondant à des mots supprimés.

Nous avons effectivement travaillé de cette manière pendant quelque temps. Les inconvénients cités nous ont fait opter pour la création de ce lexique indépendamment du programme de traduction. Cette phase de démarrage a toutefois été très utile dans la mesure où elle a permis de dégager une liste de mots qui une fois épurée des mots indésirables, a fourni la base de notre lexique. L'introduction de mots nouveaux est faite au fur et à mesure des besoins et deviendra d'ailleurs de moins en moins fréquente avec l'accroissement du nombre des documents analysés.

Ce lexique doit permettre de coder et de décoder les mots utilisés dans la description des documents. Pour ce faire, nous avons utilisé une méthode de classement par ordre alphabétique à l'aide d'un arbre binaire (voir III. B - La méthode utilisée dans le lexique des noms de paragraphes et de composantes n'offre en effet pas de bonnes prestations ici pour deux raisons :

- Les mots sont longs (24 caractères) et la méthode précédente demande de les écrire deux fois.

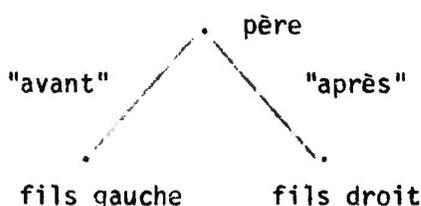
- nous pensons qu'il y aura des adjonctions de mots nouveaux relativement fréquentes (à chaque fois, il faudrait remanier la liste par ordre alphabétique).

Le lexique doit également permettre de hiérarchiser des mots. Prenons un exemple simple. Imaginons que l'on veuille sélectionner les documents qui contiennent le mot HERBIVORE. Il serait mauvais que des documents qui parlent de LAPIN ou de VACHE par exemple ne soient pas sélectionnés de façon automatique (c'est-à-dire sans que l'utilisateur précise ces mots quand il pose la question). D'où l'introduction de relations de hiérarchie entre les mots. Voir le paragraphe III. B - 3 à ce sujet.

B - 2. Méthode de classement par ordre alphabétique à l'aide d'un arbre binaire.

Un arbre binaire est une arborescence telle que chaque noeud possède au maximum deux fils appelés fils gauche et droit. Même si un noeud n'a qu'un descendant, on précise s'il est gauche ou droit.

A chaque noeud nous attacherons un mot. Un fils gauche est alphabétiquement avant son père et un fils droit est alphabétiquement après son père.



Chaque fois qu'un mot nouveau est introduit dans ce lexique, il est attaché à un noeud d'un arbre binaire suivant cette règle.

Exemple.

Pour plus de brièveté dans l'exposé, nous représenterons les mots par leur initiale (étant entendu que le classement se fait compte tenu de tous les caractères du mot).

Imaginons que les mots suivants soient introduits dans le lexique (dans l'ordre où ils sont écrits) :

D, C, F, A, B, E, H

L'arbre construit aura successivement les états représentés sur le schéma n° 3, page 3.17.

Nous donnons sur le schéma n° 4, page 3.18, l'organigramme de l'algorithme d'introduction d'un mot dans l'arbre. L'arbre est construit en itérant cet algorithme pour chaque mot nouveau.

Cette méthode revient à faire un classement par dichotomie.

Pour représenter un tel arbre en mémoire, nous utilisons une méthode classique de chaînage des lignes d'un tableau. Une ligne de ce tableau correspond à un noeud et contient :

- le mot attaché au noeud
- l'indice du fils gauche s'il existe, sinon zéro
- l'indice du fils droit s'il existe, sinon zéro.

Pour l'arbre précédent donné en exemple, on obtient :

	mot	fils gauche	fils droit
1	D	2	3
2	C	4	0
3	F	6	7
4	A	0	5
5	B	0	0
6	E	0	0
7	H	0	0

Remarques.

1. Un mot nouveau est rangé dans la première ligne libre du tableau.

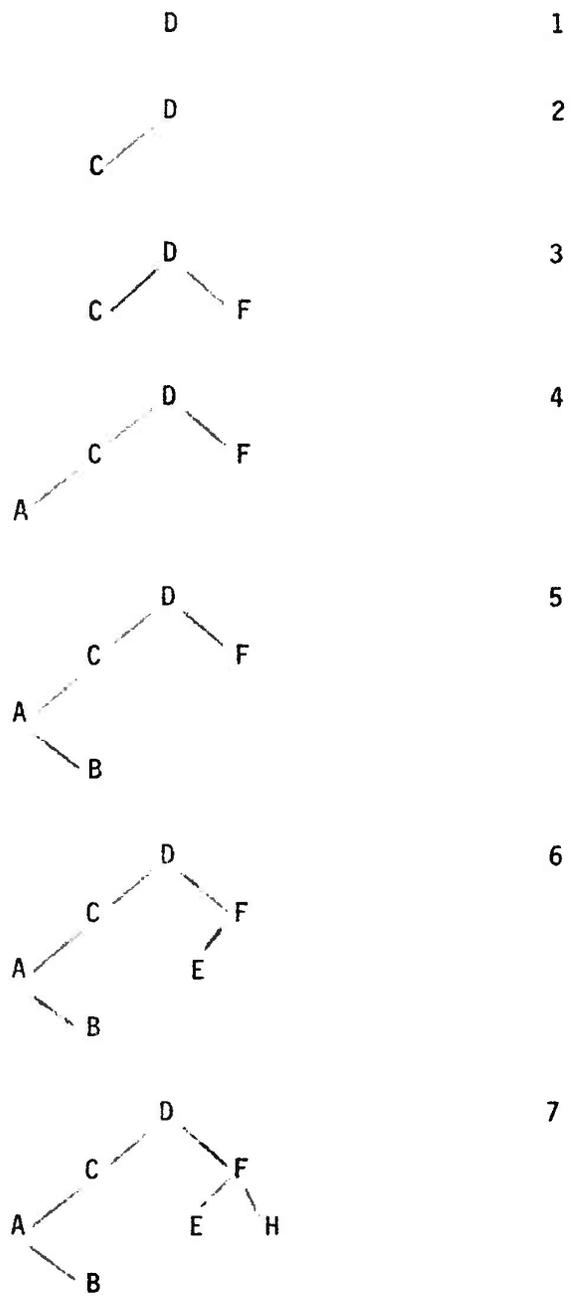
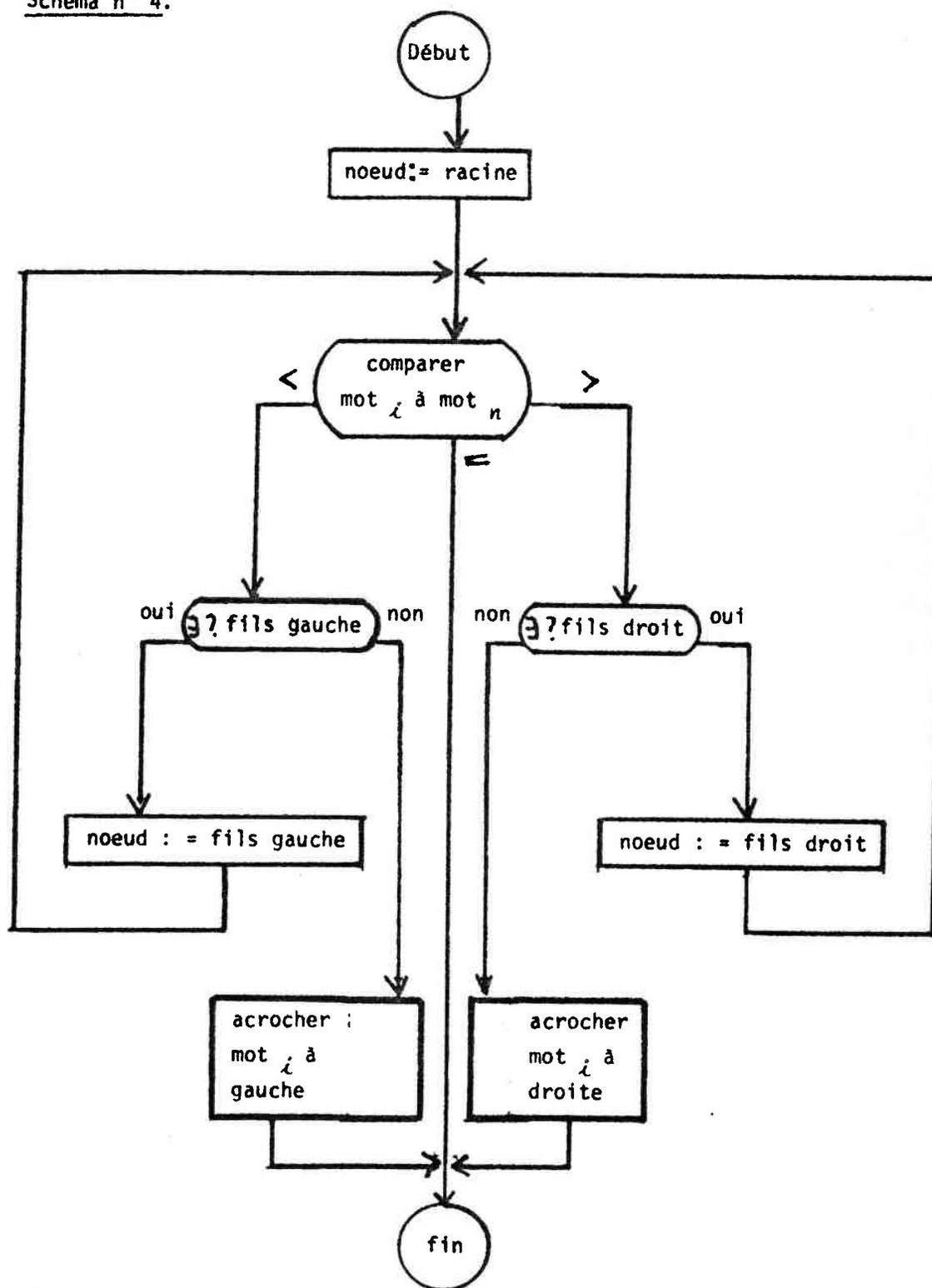
Schéma n° 3.

Schéma n° 4.



mot_i = mot à introduire
 mot_n = mot du noeud

2. Il est possible de représenter plusieurs arbres dans le même tableau à condition de savoir où se trouvent les racines (dans l'exemple la racine est implicitement représentée dans la première ligne).

Compte-tenu de ce qui précède, nous avons choisi de coder un mot par son ordre d'arrivée dans le lexique. Ainsi donc l'opération de décodage est efficace puisque le code d'un mot correspond précisément à l'indice de la ligne où il est rangé (pour l'exemple le code de F est 3).

Pour coder un mot, il suffit de le rechercher dans le tableau et de retenir le numéro de ligne où il a été découvert (s'il existe). Cette recherche se fait bien évidemment en parcourant l'arbre binaire à partir de la racine (de la même manière qu'au moment de l'introduction d'un mot) ce qui revient à faire une recherche dichotomique.

Mais revenons plus précisément à notre lexique des mots.

Nous avons classé les mots par composante afin d'améliorer les performances de la mise-à-jour et du codage. Ainsi donc pour rechercher un mot, il suffit de consulter ceux qui sont attachés à la composante à laquelle il appartient lui-même (inutile d'aller rechercher un nom d'auteur parmi les espèces d'animaux). Pour chaque composante les mots qui y sont attachés sont classés par ordre alphabétique à l'aide d'un arbre binaire, la représentation de ces arbres étant faite dans le même tableau.

Un autre avantage de ce classement des mots par composante apparaîtra plus loin pour les problèmes de hiérarchie.

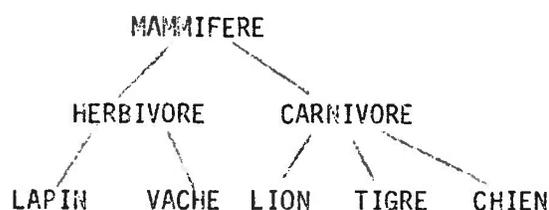
L'inconvénient est que si un mot se trouve dans plusieurs composantes, il figure plusieurs fois (avec un code différent) dans le lexique. Mais de tels mots sont rares et cet inconvénient ne nous a pas paru d'importance en comparaison du gain d'efficacité obtenu pour chaque recherche de mots (et aussi pour chaque introduction de mots).

B - 3. Relation de hiérarchie.

Ce problème a été posé dans l'introduction. Pour y répondre, nous avons introduit la possibilité d'exprimer qu'un mot est un cas particulier d'un autre mot du lexique. Par exemple : le mot HERBIVORE est un cas particulier de MAMMIFERE. (Pour faciliter la lecture de l'exposé, nous donnons des exemples de mots qui n'existent pas en réalité dans le lexique).

Nous imposons toutefois qu'un mot ne soit le cas particulier que d'un seul autre mot (structure d'arborescence en termes de graphe).

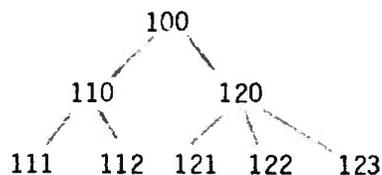
Exemple.



Une telle arborescence ne peut être définie que pour des mots qui sont attachés à la même composante (voir fin de B - 2). En outre, il ne peut exister qu'une seule arborescence de hiérarchie au plus par composante. Une arborescence peut ne hiérarchiser que certains mots parmi ceux qui sont attachés à une composante (tous les mots ne le sont pas obligatoirement).

Cette organisation pose certains problèmes dans la mesure où un mot ne peut pas être le cas particulier de plusieurs mots. Cet inconvénient est nettement atténué dans la mesure où l'on construit une arborescence par composante et que chaque mot est vu sous un seul éclairage.

Reste à adopter une représentation en mémoire d'une telle arborescence. Une méthode classique en documentation consiste à attribuer à chaque mot un code qui rend compte de sa position dans l'arborescence. Pour l'exemple précédent, les codes pourraient être :



Cette méthode offre l'avantage de ne pas avoir à représenter l'arborescence par ses liens. Nous lui avons trouvé trois inconvénients :

1. Le nombre des descendants d'un mot est limité à 9. Ce n'est pas grave dans la mesure où l'on peut faire le même codage dans un système de numération de base supérieure à 10 (En base N on a N-1 descendants possibles).

2. Ce codage ne permet plus d'accéder directement à un mot par son code comme nous le faisons (voir B - 2) car il y a des "trous".

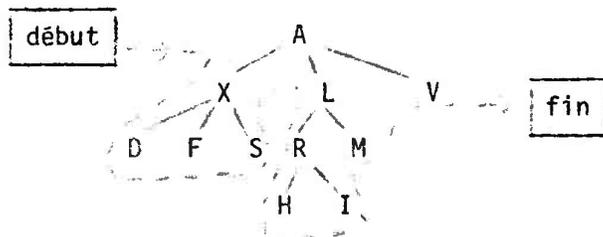
3. Nous abordons l'inconvénient le plus sérieux. La suppression ou l'adjonction de mots dans l'arborescence peut modifier certains codes déjà existants. Et par conséquent, il faut dans de telles circonstances faire un transcodage sur les documents déjà entrés en machine. A priori on peut penser que cet inconvénient s'atténue avec le temps car le lexique se stabilise. Mais par contre, le transcodage devient plus long car le nombre de documents enregistrés en machine quant à lui augmente.

Cette méthode est bonne quand le lexique existe dès la mise en route du système informatique de documentation car les modifications sont alors rares (par exemple automatiser un système existant et par conséquent bien rodé).

Nous avons préféré utiliser le codage déjà défini et choisi une autre méthode de représentation des arborescences de hiérarchie.

Notre but est de représenter une arborescence de façon à ce qu'à partir d'un mot M, nous ayons accès rapidement à tous les descendants de M. La représentation préfixée répond à ce critère. Elle s'obtient en énumérant les noeuds dans l'ordre du passage préfixé (passer par le père avant de passer par les fils) et en associant à chaque noeud son demi-degré extérieur (c'est-à-dire le nombre de ses descendants).

Exemple : Les pointillés représentent l'ordre de parcours de l'arborescence.



Représentation préfixée.

noeud :	A	X	D	F	S	L	R	H	I	M	V
degré :	3	3	0	0	0	2	2	0	0	0	0

En réalité ce ne sont pas les mots qui figurent dans le tableau mais leur code.

Rappelons que les mots sont classés par ordre alphabétique à l'aide d'un arbre binaire (voir B - 2). Si un mot fait partie d'une arborescence de hiérarchie et s'il possède des descendants, nous lui associons un pointeur vers sa position dans l'arborescence. Ceci afin de pouvoir disposer rapidement de tous les descendants d'un mot. Nous avons représenté les pointeurs par des pointillés dans le schéma ci-dessous.

Arbre binaire

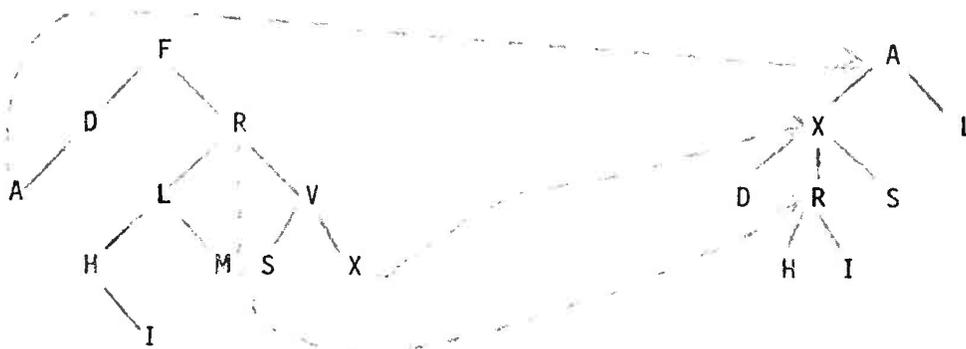
(ordre alphabétique)

ordre d'arrivée :

F, D, R, L, H, I, M, A, V, S, X

Arborescence

(hiérarchie)



La représentation de ces pointeurs est faite dans le tableau qui répertorie les mots.

Nous venons de voir les méthodes que nous utilisons pour la constitution du lexique des mots.

Ce lexique est conservé sur deux fichiers :

- le fichier des mots (FLM) contient la représentation des arbres binaires,
- le fichier hiérarchique (FH) contient la représentation des arborescences de hiérarchie.

Nous avons scindé ce lexique en deux fichiers car au moment de la traduction des documents, seule la première partie nous intéresse. Pour l'interrogation, les deux fichiers sont utilisés afin de répondre aux questions, compte tenu des structures hiérarchiques. Nous décrivons maintenant ces deux fichiers.

B - 4. Description du fichier FLM.

Ce fichier est séquentiel et conservé sur bande magnétique et il est amené intégralement en mémoire centrale au début d'un travail qui y fait référence (ceci afin d'éviter de nombreux échanges entre périphérique et mémoire centrale).

Avant de donner la description de ses enregistrements logiques, nous décrivons la manière dont il est rangé en mémoire centrale après lecture.

Tableau TLM (Tableau Lexique des Mots)

Ce tableau permet de ranger les arbres binaires de classement des mots par ordre alphabétique. Chaque ligne a le dessin suivant :

	LM	AVANT	APRES	HIER
ligne n° j :	<div style="border: 1px solid black; width: 100%; height: 20px; display: flex; justify-content: space-between; align-items: center;"> </div>			
	X(24)	COMP	COMP	COMP

- LM contient le mot cadré à gauche et complété par des blancs à droite s'il fait moins de 24 caractères.

- AVANT et APRES contiennent respectivement l'adresse (c'est-à-dire un indice de ce même tableau) des fils gauche et droit (zéro pour un fils inexistant).

- HIER (hiérarchie) est un pointeur vers l'endroit où se trouve le mot dans la représentation préfixée de l'arborescence de hiérarchie. HIER contient :

- zéro si le mot n'appartient à aucune hiérarchie ou si le mot est une feuille de l'arborescence (c'est-à-dire s'il n'a pas de descendants).

- l'adresse de la position du mot dans la représentation préfixée de l'arborescence si le mot est un noeud non terminal (c'est-à-dire s'il possède au moins un descendant).

Tableau TTETE.

Ce tableau a autant de lignes que de composantes prévues (actuellement 300).

ligne n° i :

TETE
CØMP

TETE contient l'indice de la ligne du tableau TLM où est rangée la racine de l'arbre binaire de classement par ordre alphabétique des mots attachés à la composante dont le code est i.

RTLTM (Remplissage du tableau TLM)

contient l'indice de la dernière ligne occupée de TLM (sous forme d'entier CØMP).

Enregistrements logiques de FLM.

Le premier enregistrement contient le tableau TTETE et le renseignement RTLTM soit $300 \times 4 + 4 = 1204$ octets.

Les autres enregistrements contiennent chacun 33 lignes du tableau TLM soit $(24 + 3 \times 4) \times 33 = 1188$ octets.

La taille du bloc est de 1208 octets.

Lecture de FLM.

La lecture de ce fichier est faite par un appel du module LECFLM (ordre PERFORM de COBOL). Ce module lit le fichier et le range en mémoire centrale de la manière décrite ci-dessus.

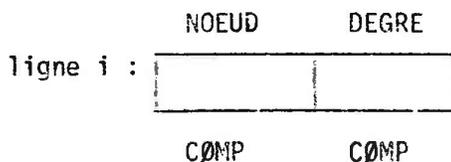
B - 5. Description du fichier FH.

Ce fichier est séquentiel et conservé sur bande magnétique. Il est amené intégralement en M.C. (mémoire centrale) au début d'un travail qui y fait référence.

Après lecture, il se présente de la manière suivante en MC :

Tableau TARBO (Tableau Arborescences).

Ce tableau permet de ranger les arborescences sous leur forme préfixée (voir B - 3). Une ligne de ce tableau a le dessin suivant :



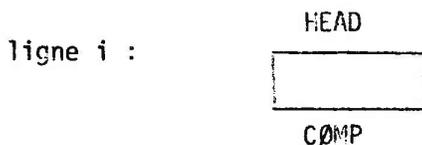
- NOEUD contient le code d'un mot (et non pas le mot lui-même comme dans l'exemple B - 3).

- DEGRE contient le nombre de fils du noeud.

Les différentes arborescences sont rangées de façon séquentielle dans ce tableau. Leur racine est repérée grâce au tableau suivant.

Tableau THEAD.

Ce tableau a autant de lignes que de composantes prévues.



HEAD contient l'indice de la ligne du tableau TARBO où est rangée la racine de l'arborescence de hiérarchie des mots de la composante de code i.

*TARBØ (Remplissage de TARBO).

contient l'indice de la dernière ligne occupée par les arborescences dans TARBO (entier CØMP).

Enregistrement de FH

Le premier enregistrement de FH contient le tableau THEAD et le renseignement RTARBO soit $300 \times 4 + 4 = 1204$ octets.

Les autres enregistrements contiennent chacun 150 lignes du tableau TARBO soit $150 \times 8 = 1200$ octets.

La taille d'un bloc est de 1208 octets.

Lecture de FH

La lecture de FH est faite par appel du module LECFH (ordre PERFORM). Ce module lit le fichier et le range en M.C. de la manière décrite ci-dessus.

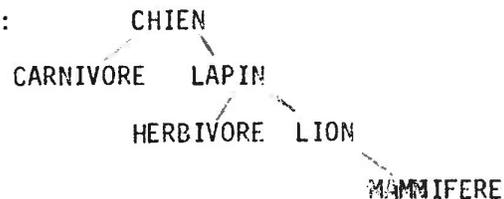
B - 6. Exemple de rangement de mots et de leur structure hiérarchique

Nous donnons ci-dessous une liste de mots avec leur rang d'arrivée dans le lexique. Pour chacun d'eux, nous précisons la composante à laquelle il appartient.

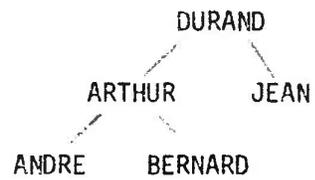
1	CHIEN	SPEC
2	DURAND	MAIN
3	ARTHUR	MAIN
4	CARNIVORE	SPEC
8	LAPIN	SPEC
9	HERBIVORE	SPEC
10	JEAN	MAIN
16	BERNARD	MAIN
17	LION	SPEC
18	MAMMIFERE	SPEC
19	ANDRE	MAIN

Les arbres binaires de classement par ordre alphabétique sont :

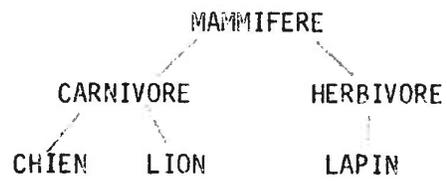
pour SPEC :



pour MAIN :

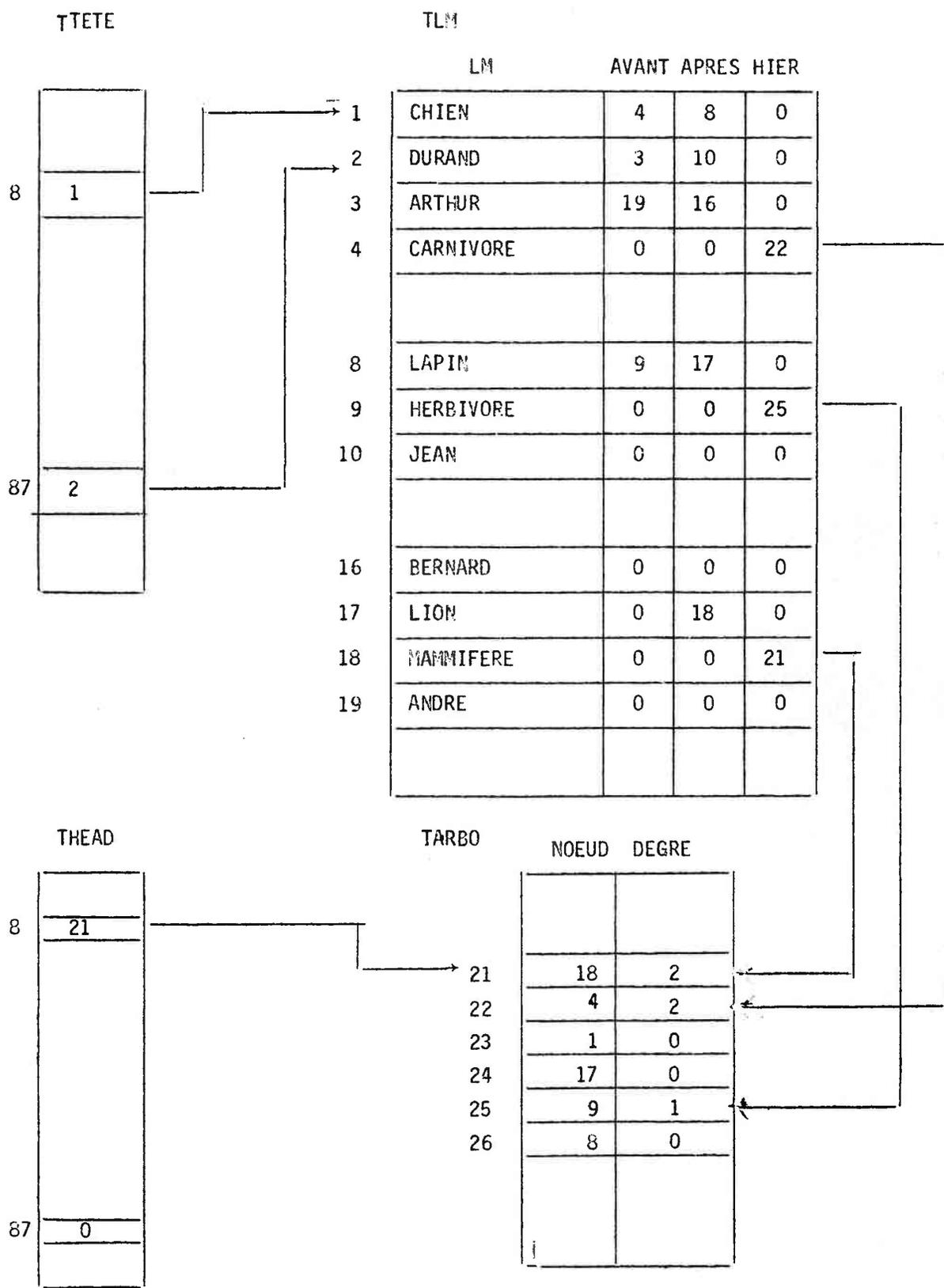


D'autre part les mots de la composante SPEC ont une structure hiérarchisée :



Nous allons représenter le contenu des différents tableaux dont nous avons parlé (après lecture des fichiers FLM et FH).

Le code de la composante SPEC est 8, celui de MAIN est 87.



B - 7. Création et mise à jour du fichier FLMB - 7. 1 - Création de FLM.

La création du lexique des mots est faite à partir de cartes perforées par un programme de nom CLM. Ce programme peut aussi servir à reconstituer le fichier en cas de destruction accidentelle.

L'utilisateur dresse la liste des mots à entrer dans le lexique sur des bordereaux en donnant pour chacun d'eux :

- son code (c'est-à-dire son rang d'apparition)
- le nom de la composante à laquelle il doit être attaché.

On utilisera une ligne du bordereau par mot avec le format suivant :

1	5
code	nom de composante / mot /

- le code est sur 4 chiffres à partir de la colonne 1
- le nom de composante est suivi d'un slash (les "blancs" sont non significatifs).
- le mot est suivi d'un slash (les "blancs" sont non significatifs)
- le code sert également de numérotage par ordre croissant des lignes.

L'attribution n'est pas faite automatiquement pour éviter toute erreur en cas de reconstitution du fichier après destruction accidentelle. Il serait en effet catastrophique d'attribuer un autre code à un mot, les documents dans leur forme interne perdant alors toute signification (cette forme contient le code des mots). Le programme signale les erreurs de séquence dans les cartes.

Après création, les cartes sont conservées dans des archives.

B - 7. 2 - Mise à jour du fichier FLM.

Le programme de mise à jour du fichier FLM permet d'effectuer deux types d'opérations :

- introduire des mots dans le lexique après sa création (B - 7. 1)
- remplacer un mot par un autre mot.

1. Introduire un mot dans le lexique.

Cette introduction se fait à partir de cartes dont le format est le même que lors de la création du lexique (voir B - 7. 1). En particulier, le code attribué au mot doit être correct (code non encore utilisé dans le lexique).

Plusieurs mots peuvent être introduits en même temps dans le lexique par le programme. Les cartes correspondantes sont mises ensuite dans les archives des mots.

2. Remplacer un mot par un autre.

Le cas se présente quand un mot a été introduit dans le lexique sous une orthographe incorrecte ou si l'on veut remplacer un mot par un synonyme que l'on juge meilleur que le mot initial. Pour une telle opération le code du mot nouveau est le même que celui de l'ancien et par conséquent la forme interne des documents n'a pas à être modifiée.

Le format des cartes qui permettent cette opération est le même que celui de l'introduction de mots mais le dernier slash doit être suivi de la lettre R.

Exemple.

Le mot HEMORHAGE a été introduit dans le lexique sous le code 101 et pour la composante COND (condition). On veut le remplacer par HEMORRHAGE (orthographe correcte).

On écrira :

0101	COND / HEMORRHAGE / R
------	-----------------------

Après l'opération de remplacement, les cartes de ce type vont remplacer dans les archives des mots les anciennes versions de ceux-ci. Ainsi la reconstruction éventuelle du fichier des mots après destruction accidentelle fait intervenir la bonne version d'un mot (Le programme de création ne tient pas compte des indications qui suivent le deuxième slash du format d'entrée).

Remarques.

1. Lorsqu'une opération de remplacement est entreprise, il y a bien sûr réorganisation de l'arbre binaire qui rend compte de l'ordre alphabétique des mots (le mot nouveau n'occupe pas en général la même place dans l'arbre que le mot remplacé). Les mots ne sont pas déplacés par ce remaniement, seuls des pointeurs sont modifiés et par conséquent aucun code n'est modifié.
2. Nous avons interdit la suppression pure et simple d'un mot du lexique pour éviter que des documents puissent contenir des codes qui ne correspondent plus à rien.

B - 8. Création et mise à jour du fichier FH.B - 8. 1 - Création.

Le programme de création (CFH) de ce fichier ne fait qu'initialiser le tableau `THEAD` et `RTARBO` à zéro (premier enregistrement). Les arborescences sont construites à l'aide du programme de mise à jour (MFH).

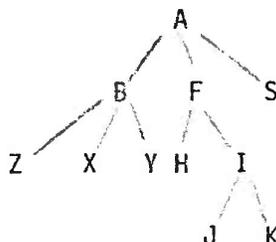
B - 8. 2 - Mise à jour.

Les opérations permises sont de deux types :

- a) création d'une arborescence de hiérarchie
- b) modifier une arborescence déjà existante.

Nous allons d'abord donner une méthode de représentation "linéaire" d'une arborescence nécessaire pour son entrée en machine.

- Les mots sont énumérés dans l'ordre du passage préfixé
- Une famille de mots (c'est-à-dire l'ensemble des mots ayant le même père) est écrite entre parenthèses.
- Les mots d'une même famille sont séparés par des virgules.

Exemple.

Description : A (B (Z, X, Y), F (H, I (J, K)), S)

Les opérations de mise à jour sont effectuées à partir de cartes dont nous allons donner le format général. Puis nous expliciterons ces opérations.

Format général.

La demande d'exécution d'une opération sur une arborescence peut nécessiter plusieurs lignes du bordereau.

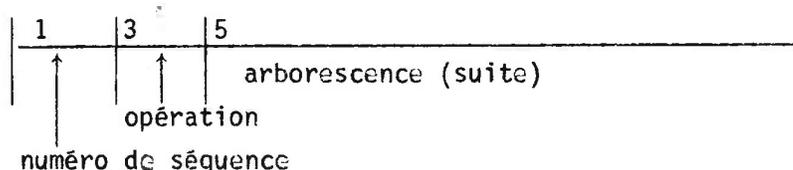
Le format de la première ligne est le suivant :

1	3	5	
0	0		nom de composante / arborescence

↑
opération

- Les colonnes 3 et 4 précisent la nature de l'opération à effectuer.
- Le nom de composante concerné par l'opération est écrit à partir de la colonne 5 et est suivi d'un slash (les blancs sont non significatifs).

Si une ligne ne suffit pas pour décrire l'arborescence, on continue sur les lignes suivantes en les numérotant par ordre croissant et en répétant le nom de l'opération dans les colonnes 3 et 4. Les blancs sont ignorés dans la partie qui concerne la description des arborescences. Le format d'une telle ligne est :



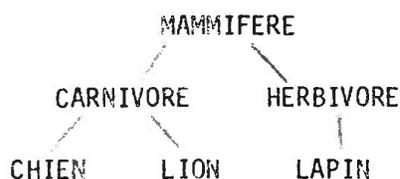
a) Création d'une arborescence.

Il s'agit, pour une composante, de créer une structure hiérarchique entre certains mots. Cette opération ne peut être entreprise que s'il n'existe pas encore d'arborescence pour la composante en question (nous avons en effet imposé qu'il n'y ait qu'une seule arborescence par composante, se reporter en III. B - 3).

Les colonnes 3 et 4 contiennent alors les caractères c :

Exemple.

Création de la structure suivante pour la composante SPEC (specious)



On écrira :

0 0	C:	SPEC / MAMMIFERE (CARNIVORE (CHIEN, LION),
0 1	C:	HERBIVORE (LAPIN))

Il est bien sûr nécessaire que les mots cités dans une arborescence que l'on veut créer existent déjà dans le lexique des mots. Sinon le programme signale une erreur.

On peut être amené à entrer dans le lexique des mots qui ne sont pas utilisés dans les documents mais qui servent à construire ces structures. Dans notre exemple, le mot MAMMIFERE peut ne pas apparaître dans les documents car il est trop général. Mais à l'interrogation, il permet de sélectionner les documents qui contiennent un de ses descendants.

b) Modifications d'une arborescence.

Elles sont de deux types :

- compléter une arborescence par une sous-arborescence sans perturber l'existant.
- supprimer les descendants d'un mot d'une arborescence.

b.1) Compléter une arborescence sans perturber l'existant.

Cette opération permet d'accrocher une sous-arborescence à un mot qui fait déjà partie d'une arborescence, ce mot pouvant être à n'importe quel niveau. "Sans perturber l'existant" signifie qu'aucun lien de l'arborescence à compléter ne doit être détruit.

Règles à observer :

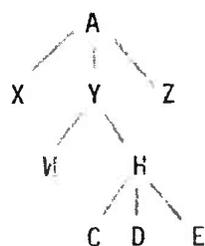
- la racine de la sous-arborescence doit appartenir à l'arborescence existante.
- tous les autres mots de la sous-arborescence ne doivent pas encore appartenir à l'arborescence.

Si ces règles ne sont pas respectées, le programme signale une erreur

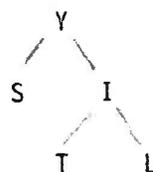
La demande d'exécution d'une telle opération se fait à l'aide du format général déjà défini avec les caractères A: (adjoindre) dans les colonnes 3 et 4.

Exemple.

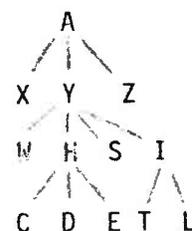
arborescence existante



sous-arborescence à introduire



arborescence après l'opération



Pour obtenir ce résultat, il aura suffit d'écrire :

0 0	A:	C1 / Y (S, I (T, L))
-----	----	----------------------

On a supposé que les mots sont attachés à la composante C1.

Exemple 2.

Un cas particulier consiste à introduire un seul mot dans une arborescence :

0 0	A:	C1 / Z (F)
-----	----	------------

Le mot F va être accroché au mot Z de l'arborescence précédente.

Exemple 3.

La demande suivante :

0 0	A:	C1 / H (S)
-----	----	------------

est erronée car S appartient déjà à l'arborescence.

b.2) Supprimer les descendants d'un mot d'une arborescence.

Cette opération a été introduite pour permettre la destruction de liens hiérarchiques.

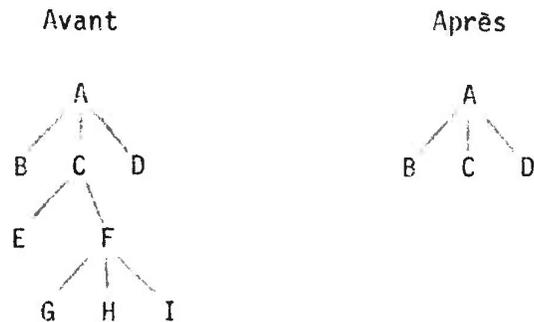
La zone opération du format général contient S:

La zone arborescence du format contient le mot dont on veut supprimer les descendants (nous entendons par descendants les mots de la sous-arborescence qui a pour racine le mot en question).

Exemple.

Supprimer les descendants du mot C dans l'arborescence de la composante C2.

0 0	S:	C2 / C
-----	----	--------



Les liens hiérarchiques sont supprimés, mais les mots eux-mêmes restent bien évidemment toujours dans le lexique.

Pour demander la destruction pure et simple d'une arborescence, on laisse à blanc la zone de description d'arborescence du format général.

Exemple.

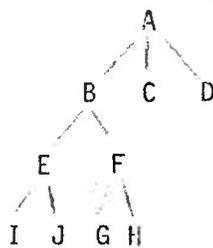
Supprimer la structure hiérarchique des mots de la composante C3 :

0	0	S:	C3/
---	---	----	-----

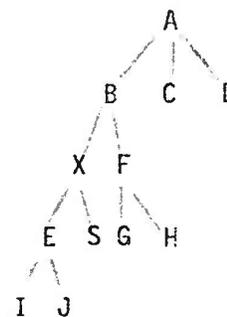
Les deux opérations précédentes (A: et S:) permettent de remanier complètement une arborescence.

Exemple :

arborescence existante



on veut obtenir :



On écrira (on suppose que la composante C4 est concernée) :

0	0	S:	C4 / B
0	0	A:	C4 / B (X (E (I, J), S), F (G, H))

B - 8. 3 - Remarque.

Les opérations sur les arborescences de hiérarchie mettent également en cause le fichier des mots FLM. Rappelons en effet que dans le tableau TLM qui répertorie les mots, nous avons introduit un pointeur (HIER) vers la représentation des arborescences. Le programme de mise à jour du fichier FM modifie aussi le fichier FLM. Nous tenions à le signaler.

C - Les fichiers des documents (FDOC et FDOCD).

C - 1. Généralités.

Le programme de traduction TRD prend en compte les documents perforés sur cartes (FCARTE) et les traduit dans une forme que nous avons baptisée "interne".

La forme interne a été choisie de façon à satisfaire deux critères :

- permettre une consultation rapide d'un document (le but final de l'entreprise est d'interroger ces documents).
- occuper une place mémoire de taille raisonnable.

En cas de conflit entre ces deux critères, nous avons donné la priorité au premier, estimant que la vitesse de consultation était essentielle.

Notons qu'une fois entré en machine, un document n'a plus à être modifié. Cette remarque a également influé sur le choix de la forme interne. En particulier, le code d'un mot reste toujours le même, ce qui n'est pas vrai si l'on choisit de faire un codage qui rend compte de la structure hiérarchique du lexique.

Nous allons expliquer d'abord en C-2 la façon dont sont rangés les listes de paragraphes dans des tableaux au cours de la traduction d'un document.

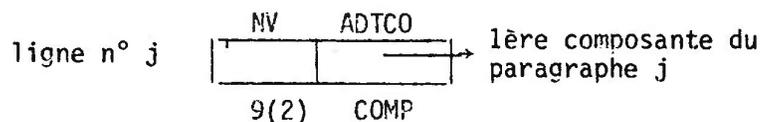
Nous décrivons ensuite en C-3 les fichiers FDOC et FDOCD dont la partie essentielle des enregistrements est constituée par regroupement des tableaux dont nous parlons en C-2.

C - 2. Forme interne des listes de paragraphes.

Au cours de la traduction des documents, le programme TRD construit des tableaux.

C - 2. 1 - Tableau des paragraphes (TP)

La ligne n° j de ce tableau concerne le paragraphe dont le nom est codé par j.

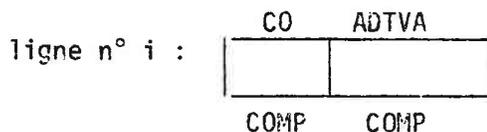


- NV : nombre de vecteurs du paragraphe j.
- ADTCO : adresse de la première composante du paragraphe j (les composantes sont rangées dans le tableau TCO). Cette zone est à zéro si le paragraphe j n'est pas présent dans le document.

Ce tableau possède 50 lignes (nous prévoyons que le nombre de paragraphes n'excèdera pas ce nombre). Ce tableau permet de détecter immédiatement si le paragraphe de code k est présent dans le document (par un test sur ADTCO) et si oui accéder à son premier vecteur (important pour les performances de l'interrogation).

C - 2. 2 - Tableau des composantes (TCO)

Chaque ligne de ce tableau concerne une composante.



- CO : contient le code d'un nom de composante.
- ADTVA : contient un nombre entier dont la valeur absolue est l'adresse de la première valeur de la liste de valeurs de la composante dont le code est CO (les valeurs sont rangées dans un tableau TVA).

Les composantes d'un même vecteur sont rangées séquentiellement dans le tableau TCO. La fin d'un vecteur est marquée en rendant négative la valeur contenue dans ADTVA.

Exemple.

Le paragraphe de code 36 est constitué de deux vecteurs. Le premier vecteur a deux composantes de codes 60 et 64, le deuxième vecteur a trois composantes de codes 1, 7, 12.

Tableau TP

	NV	ADTCO
36	2	i

Tableau TCO

	CO	ADTVA
	60	k_1
	64	$-k_2$
	1	k_3
	7	k_4
	12	$-k_5$

Les k_j ($k_j > 0$) sont des indices du tableau des valeurs TVA que nous allons étudier maintenant.

C. 2. 3. Tableau des valeurs (TVA):

Ce tableau sert à ranger les listes de valeurs (suite de valeurs élémentaires adjectivées ou non).

L'accès à ces listes de valeurs se fait par le contenu de ADTVA donné dans le tableau TCO qui vient d'être décrit.

Une ligne de ce tableau a le dessin suivant :

VE/VR	ADJ	MN	IND
COMP/COMP-1	COMP	X	X

Zone VE/VR. Cette zone contient :

- soit le code d'un mot
- soit un nombre

Cette zone est désignée par VE (respt VR) si son contenu est entier (respt réel)

Zone MN. Cette zone contient un caractère qui précise le contenu de VE/VR.

M : le contenu de VE/VR est le code d'un mot (c'est-à-dire un nombre entier, rappelons le)

R : le contenu de VE/VR est un nombre réel

N : le contenu de VE/VR est un nombre entier.

Les valeurs standards (voir carte syntaxique en II. 2 - B) sont rangées de manière particulière. La zone VE/VR n'est pas utilisée, seule la zone MN contient l'un des caractères suivants :

A	pour la valeur standard	A	(augmentation)
D	" " "	D	(diminution)
P	" " "	P	(probabilité)
C	" " "	PC	(pourcentage)
B	" " "	NS	(non significatif)
E	" " "	E	(étudié)
I	" " "	I	(indéfini)

Pour ces valeurs le contenu de VE/VR est indifférent.

Zone IND.

Le rangement d'une fourchette nécessite une ou deux lignes du tableau TVA suivant le nombre de ses termes. La zone IND est un indicateur qui contient l'un des caractères suivants :

- F VE/VR contient le premier terme d'une fourchette à deux termes.
 - T VE/VR contient le deuxième terme d'une fourchette à deux termes.
 - S VE/VR contient le terme d'une fourchette de type supérieur à.
 - I VE/VR contient le terme d'une fourchette de type inférieur à.
- "blanc"^{on}/ne se trouve dans aucun des cas précédents.

Zone ADJ.

Cette zone contient un nombre entier dont la valeur absolue est :

- 1000 si le contenu de VE/VR n'est pas adjectivé.
- l'adresse du premier adjectif de la liste si le contenu de UE/UR est adjectivé.

Le tableau TVA est subdivisé en 3 tranches. La première contient des listes de valeurs élémentaires, la deuxième des listes d'adjectifs de niveau 1 et la troisième des listes d'adjectifs de niveau 2. Les valeurs élémentaires sont rangées de façon séquentielle dans TVA. Pour marquer la fin d'une liste la zone ADJ de la dernière valeur est rendue négative.

Nous allons donner un exemple de rangement de composantes.

$C1 = 1, <2, >3, <1:5> (10.2), H (B, C, L),$

$X (W (<12 E + 3 : 15 E + 3> , -36), Z (Y, W)) ;$

où $C1$ est un nom de composante et les lettres des initiales de mots. Dans le tableau TVA nous ne remplaçons pas les mots par leur code.

CO	ADTVA	VE/VR	ADJ	MN	IND	
C1	s					
		1	1000	N	b	listes de valeurs élémentaires
		2	1000	N	I	
		3	1000	N	S	
		1	1000	N	F	
		5	k	N	T	
		H	k+1	M	b	
		X	-(k+4)	M	b	
k	10.2	-1000	R	b	listes d'adjectifs de niveau 1	
k+1	B	1000	M	b		
	C	1000	M	b		
	L	-1000	M	b		
k+4	W	I	M	b		
	Z	-(1+3)	M	b		
1	12000	1000	R	F	listes d'adjectifs de niveau 2	
	15000	1000	R	T		
	-36	-1000	N	b		
1+3	Y	1000	M	b		
	W	-1000	M	b		

C - 2.4 - Tableau des expériences (TEX)

Rappelons rapidement que l'on peut attacher un numéro d'expérience à l'aide de la composante EXP. Etant donné un vecteur on sait donc à quelle expérience il appartient. Mais il est utile de connaître au moment de l'interrogation la liste des vecteurs concernés par une expérience. Au fur et à mesure de la traduction sous forme interne, le programme va dresser cette liste pour chaque numéro.

Ces listes sont rangées dans le tableau TEX dont chaque ligne est découpée comme suit :

EXPAR	ADVEC	NEXP
COMP	COMP	9(2)

Zone NEXP contient le n° de l'expérience.

Zone ADVEC contient l'adresse du vecteur concerné par l'expérience (c'est-à-dire un n° de ligne du tableau TCO)

Zone EXPAR contient le code du nom du paragraphe auquel appartient le vecteur (l'accès au vecteur par ADVEC ne le donne pas).

Les lignes de ce tableau sont triées par numéros d'expérience croissant.

Donnons un exemple :

PATT = SPEC = DOG ; NBR = 6 ; EXP = 1 #

SPEC = CALF ; EXP = 5,8 /

PUMP = TYPE = X ; EXP = 1,5 #

TYPE = Y ; EXP = 8 /

Le code de PATT est 1, celui de PUMP est 7.

		<u>TP</u>	
		NV	ADTCO
1		2	1
7		2	6

		<u>TCO</u>	
		CO	ADTVA
1	SPEC	1	
	NBR	2	
	EXP	-3	
4	SPEC	4	
	EXP	-5	
6	TYPE	7	
	EXP	-8	
8	TYPE	10	
	EXP	-11	

		<u>TVA</u>			
		VE/VR	ADJ	MN	IND
1	DOG	-1000		M	b
2	6	-1000		N	b
3	1	-1000		N	b
4	CALF	-1000		M	b
5	5	1000		N	b
6	8	-1000		N	b
7	X	-1000		M	b
8	1	1000		N	b
9	5	-1000		N	b
10	Y	-1000		M	b
11	8	-1000		N	b

TEX

EXPAR	ADVEC	NEXP
1	1	1
7	6	1
1	4	5
7	6	5
1	4	8
7	8	8

C - 2. 5 - Remarque à propos de la composante MODIF.

Nous avons déjà signalé qu'après l'entrée d'un document en machine, il ne subsistait plus de traces de la composante MODIF (voir les remarques de II. 3 - E).

Le lecteur aura deviné que lorsque nous disions que les composantes sont recopiées cela consiste pour la forme interne à recopier uniquement des lignes du tableau TCO (Deux lignes de ce tableau pointent alors vers la même liste des valeurs).

Lorsque le programme "recopie" la composante EXP, il met à jour le tableau TEX. En effet, le vecteur résultant est alors attaché à une ou plusieurs expériences.

Exemple.

$P1 = C1 = 1, 2 ; C2 = 3, 4 (5, 6, 7) \neq$

MODIF = 1 ; C3 = 7, 8 /

IP		TCO		TVA																																																																																			
<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">NV</th> <th style="border: none;">ADTCO</th> </tr> </thead> <tbody> <tr> <td style="border: 1px solid black; text-align: center;">2</td> <td style="border: 1px solid black; text-align: center;">s</td> </tr> </tbody> </table>	NV	ADTCO	2	s	s	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;">CO</th> <th style="border: none;">ADTVA</th> </tr> </thead> <tbody> <tr> <td style="border: 1px solid black; text-align: center;">C1</td> <td style="border: 1px solid black; text-align: center;">1</td> </tr> <tr> <td style="border: 1px solid black; text-align: center;">C2</td> <td style="border: 1px solid black; text-align: center;">-(1+2)</td> </tr> <tr> <td style="border: 1px solid black; text-align: center;">C1</td> <td style="border: 1px solid black; text-align: center;">1</td> </tr> <tr> <td style="border: 1px solid black; text-align: center;">C2</td> <td style="border: 1px solid black; text-align: center;">1+2</td> </tr> <tr> <td style="border: 1px solid black; text-align: center;">C3</td> <td style="border: 1px solid black; text-align: center;">-(1+3)</td> </tr> <tr> <td style="border: 1px solid black; height: 20px;"></td> <td style="border: 1px solid black;"></td> </tr> </tbody> </table>	CO	ADTVA	C1	1	C2	-(1+2)	C1	1	C2	1+2	C3	-(1+3)			1 1+2 1+3 k	<table style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="border: none;"></th> <th style="border: none;">UE/UR</th> <th style="border: none;">ADJ</th> <th style="border: none;">MN</th> <th style="border: none;">IND</th> </tr> </thead> <tbody> <tr> <td style="border: none;"></td> <td style="border: 1px solid black;"></td> </tr> <tr> <td style="border: none; text-align: center;">1</td> <td style="border: 1px solid black; text-align: center;">1</td> <td style="border: 1px solid black; text-align: center;">1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; text-align: center;">2</td> <td style="border: 1px solid black; text-align: center;">-1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none; text-align: center;">1+2</td> <td style="border: 1px solid black; text-align: center;">3</td> <td style="border: 1px solid black; text-align: center;">1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; text-align: center;">4</td> <td style="border: 1px solid black; text-align: center;">-k</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none; text-align: center;">1+3</td> <td style="border: 1px solid black; text-align: center;">7</td> <td style="border: 1px solid black; text-align: center;">1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; text-align: center;">8</td> <td style="border: 1px solid black; text-align: center;">-1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black;"></td> </tr> <tr> <td style="border: none; text-align: center;">k</td> <td style="border: 1px solid black; text-align: center;">5</td> <td style="border: 1px solid black; text-align: center;">1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; text-align: center;">6</td> <td style="border: 1px solid black; text-align: center;">1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black; text-align: center;">7</td> <td style="border: 1px solid black; text-align: center;">-1000</td> <td style="border: 1px solid black; text-align: center;">N</td> <td style="border: 1px solid black; text-align: center;">b</td> </tr> <tr> <td style="border: none;"></td> <td style="border: 1px solid black;"></td> </tr> </tbody> </table>		UE/UR	ADJ	MN	IND						1	1	1000	N	b		2	-1000	N	b	1+2	3	1000	N	b		4	-k	N	b	1+3	7	1000	N	b		8	-1000	N	b						k	5	1000	N	b		6	1000	N	b		7	-1000	N	b					
NV	ADTCO																																																																																						
2	s																																																																																						
CO	ADTVA																																																																																						
C1	1																																																																																						
C2	-(1+2)																																																																																						
C1	1																																																																																						
C2	1+2																																																																																						
C3	-(1+3)																																																																																						
	UE/UR	ADJ	MN	IND																																																																																			
1	1	1000	N	b																																																																																			
	2	-1000	N	b																																																																																			
1+2	3	1000	N	b																																																																																			
	4	-k	N	b																																																																																			
1+3	7	1000	N	b																																																																																			
	8	-1000	N	b																																																																																			
k	5	1000	N	b																																																																																			
	6	1000	N	b																																																																																			
	7	-1000	N	b																																																																																			

Nous allons passer maintenant à la description des fichiers de documents après avoir vu comment étaient rangées les listes de paragraphes.

C - 3. Description des fichiers FDOC et FDOCD.

C - 3. 1 - Fichier FDOC.

Le fichier FDOC est de type séquentiel indexé, l'indicatif d'un document faisant office de clé. Le support de ce fichier est le disque magnétique. Le fichier est exploité en accès direct pour les opérations de mise à jour :

- insertion d'un nouveau document.
- remplacement d'un document existant par une nouvelle version (voir III - 2 - B - 2).

Au moment du transfert des documents sur le fichier définitif FDOCD il est exploité en accès séquentiel.

Les enregistrements logiques sont de taille fixe et contiennent chacun la forme interne d'un document. Celle-ci étant de longueur variable, la taille de l'enregistrement a été choisie assez grande pour pouvoir les contenir (la taille fixe est exigée pour les opérations de mise à jour).

C - 3. 2 - Fichier FDOCD.

Le fichier FDOCD est un fichier de type séquentiel sur bande magnétique. Il est mis à jour périodiquement à partir de FDOC, c'est-à-dire qu'il est rallongé.

Le fichier est exploité en accès séquentiel par l'interrogation.

Les enregistrements logiques de ce fichier sont de taille variable et sont constitués de la partie utile de ceux de FDOC.

C - 3. 3 - Enregistrements logiques.

Les enregistrements logiques de FDOC et FDOCD ne diffèrent que par leur taille.

Un tel enregistrement est constitué de deux zones :

- la zone des renseignements
- la zone des tableaux

Nous verrons plus loin ce que sont les renseignements

Zone des tableaux.

Il ne subsiste plus que deux tableaux dans l'enregistrement :

- TP (tableau de paragraphes)
- TVA (tableau des valeurs).

Le tableau TP est identique à la description que nous en avons faite précédemment.

Le tableau TVA contient dans ses premières lignes le titre de l'article. Le numéro de la dernière ligne occupée par ce titre est conservée dans un mot de nom FTi.

Les lignes qui suivent contiennent les listes de valeurs élémentaire les listes d'adjectifs de niveau 1 puis celles d'adjectifs de niveau 2. (il y a "tassage" des zones décrites précédemment).

A la suite se trouvent les tableaux TCO et TEX.

Enfin le commentaire (s'il existe) est rangé dans les lignes de TVA qui suivent.

Le schéma de la page suivante montre les différentes zones du tableau TVA.

Notons que lorsque le tableau TVA est construit dans cette forme définitive, tous les pointeurs sont mis à jour. En d'autres termes, ce tableau peut être utilisé sans qu'il soit besoin de l'éclater (à l'interrogation).

Ces opérations de translation permettent de compenser les tailles relatives des différentes zones et de construire des enregistrements de longueur variable (COBOL Cii exige qu'il n'y ait qu'un seul tableau de longueur variable par enregistrement). D'où un gain de place.

Zone des renseignements.

Ils sont de deux sortes :

1. Les données de la partie identification d'un document :

CODEUR	X	n° du codeur
NUDOC	X(3)	n° du document
REF	X	référence
AN1	X(4)	année de parution
TOME	X(4)	tome ou n° de revue
PGD	X(4)	page début
PGF	X(4)	page fin

FTi	TITRE X(10)							
	VE/VR COMP/COMP-1	ADJ COMP	MN X	IND X	listes de valeurs élémentaire			
FTVA					listes d'adjectifs de niveau			
FTVA1					listes d'adjectifs de niveau 2			
FTVA2								
FTCO					<p>TCO</p> <table border="1"> <tr> <td>CO COMP</td> <td>ADTVA COMP</td> </tr> </table>	CO COMP	ADTVA COMP	
CO COMP	ADTVA COMP							
FTEX					<p>TEX</p> <table border="1"> <tr> <td>EXPAR COMP</td> <td>ADVEC COMP</td> <td>NEXP 9(2)</td> </tr> </table>	EXPAR COMP	ADVEC COMP	NEXP 9(2)
EXPAR COMP	ADVEC COMP	NEXP 9(2)						
NBLTVA	COMMENTAIRES X(10)							
	Zone inoccupée par le document et qui n'existe pas dans les enregistrements de FDOCD							

Tableau TVA

2. Les renseignements qui indiquent la fin des différentes zones de TVA :

FTi : dernière ligne occupée par le titre
FTVA : " par les listes de valeurs élémentaire
FTVA1 : " par les listes d'adjectifs de niveau1
FTVA2 : " par les listes d'adjectifs de niveau2
FTCO : " par le tableau TCO
FTEX : " par le tableau FTEX
NBLTVA: nombre total de lignes du tableau TVA.

III. 4 - REALISATION.

Les programmes ont été testés sur la machine CII 10 070 de l'IUCA pendant les sessions d'exploitation sous SIRIS-7 (multiprogrammation) et sont écrits en langage COBOL 65 CII.

Nous ne reproduisons pas ici leur description détaillée (dossier de programmation), celle-ci ne présentant d'intérêt que pour les personnes amenées à faire leur maintenance.

La liste des programmes cités au cours de ce chapitre est la suivante:

1. LPC. Création du lexique des noms de paragraphes et de composantes (FLPC). Se reporter à III.3 - A, page 3.10.
2. CLM et MLM. Création et mise à jour du lexique des mots (FLM). Se reporter à III. 3 - B, page 3.29.
3. CFH et MFH. Création et mise à jour du fichier des arborescences de hiérarchie (FH). Se reporter à III. 3 - B - 8, page 3.31.
4. TRD. Traduction^{des documents} en forme interne à partir de leur forme externe et conservation sur le fichier FDOC.

Il est intéressant de noter que ce programme fait 987 cartes dont 640 pour la partie PROCEDURE DIVISION. L'exécution de ce programme nécessite 43 pages en mémoire centrale (1 page = 512 mots de 32 bits) soit en gros 22 K mots.

Le temps d'exécution (temps d'Unité Centrale) pour traduire un document est de l'ordre de 1.2 secondes soit une cinquantaine de documents à la minute (Ce nombre a été déterminé sur un petit nombre d'exemples).

5. TRF. Transfert du fichier FDOC sur le fichier FDOC. Voir schéma n° 1, page 3

CHAPITRE IV

INTERROGATION DU FICHER DES DOCUMENTS

IV. 1 - INTRODUCTION.

Nous appelons interrogation l'exploitation que nous voulons faire du fichier des documents et questions les demandes d'exploitation.

L'interrogation est prévue pour travailler en batch, c'est-à-dire qu'un train de questions est rassemblé et traité en une seule fois. Les questions sont entrées en machine par cartes et s'adressent à l'ensemble des documents du fichier FDOCD (il est possible de donner des limites afin de n'interroger qu'une partie du fichier).

Nous avons entrepris de dépasser le stade de la sélection habituellement retenue par les systèmes documentaires, ce qui explique la description relativement riche que nous avons faite des documents (La sélection a pour but de rechercher les documents qui répondent à un ou plusieurs critères formulés dans la question, le résultat d'une sélection est la donnée des documents qui satisfont aux critères. On qualifie ces documents de pertinents). Nous voulons en outre faire des traitements sur les documents pertinents (par exemple calculs de moyennes).

Pour chaque document du fichier, une question est entièrement exécutée avant de passer au document suivant. Une question présente une succession de sélections partielles suivies éventuellement d'un traitement à effectuer. Par exemple : si le document (en cours d'étude) parle de chiens, de leur âge et de leur nombre alors calculer leur moyenne d'âge sinon abandonner le document..etc

Nous nous sommes attachés à trouver une formulation des questions qui soit facile à mettre en oeuvre par l'utilisateur. Nous avons abouti à une représentation sous forme de table de sélection et de traitement (voir IV. 2), l'inspiration étant venue des tables de décision.

Une telle table a l'avantage de dissocier la description des sélections et des traitements bien qu'il y ait imbrication à l'exécution. La partie sélection est définie à ce jour. Il reste à approfondir la définition des traitements qui n'est que partielle. Ce travail ne pourra être mené à bien que par une coopération entre utilisateurs et informaticiens comme ce fut le cas pour la mise au point du langage de description des documents. Ces recherches pourront

alors aboutir à un langage adapté aux types de traitements demandés par l'utilisateur (Nous voulons éviter la solution qui consiste à écrire ces traitements dans un langage évolué tels que COBOL ou FORTRAN, afin de donner à l'utilisateur un outil de travail adapté à ses compétences). Ces recherches ont été démarrées par M. TUETÉY.

D'autre part, nous envisageons de définir des traitements globaux (par opposition à ceux de la table que nous appelons locaux) et qui porteraient sur l'ensemble des documents pertinents pour une question. De tels traitements permettraient de faire des études statistiques par exemple.

Une question complète est constituée de cinq parties :

1. Identification de la question
2. Limites d'exploitation du fichier des documents
3. Sélection sur l'entête d'un document
4. Table de sélection et de traitement
5. Traitement global.

Les parties 1, 2, 3 seront exposées en même temps que la manière de présenter une question sur un bordereau de perforation (IV. 3).

IV. 2 - TABLE DE SELECTION ET DE TRAITEMENT.

La table de sélection et de traitement est la partie essentielle d'une question et présente quatre subdivisions disposées comme le montre le dessin ci-dessous.

Texte de la question	Ordres de sélection
Traitements	Ordres de traitement

Nous allons étudier les quatre subdivisions de cette table.

A - Subdivision "texte de la question".

A - 1. Description et règles d'écriture.

La subdivision "texte de la question" permet de décrire les informations à sélectionner (l'ordre d'exécution des sélections est donné dans la subdivision "ordres de sélection").

Le texte d'une question reprend les notions du langage de description des documents (paragraphe, vecteurs, composantes, valeurs) décrites au chapitre II et sur lesquelles nous ne revenons pas ici.

Cette subdivision de la table a trois colonnes que nous identifierons de gauche à droite par les lettres P, C et V. Chaque ligne est donc découpée en trois parties P, C et V qui contiennent respectivement un nom de paragraphe ou des blancs, un nom de composante ou des blancs, une liste de valeurs (avec la syntaxe du chapitre II) ou des blancs.

Avant de donner les règles précises d'écriture de cette subdivision nous présentons un exemple afin de fixer les idées. Les informations à sélectionner sont :

- des chiens de moins de 3 ans d'âge et qui pèsent de 10 à 20 kg.
- des veaux.

- un pompage par ballonnet intra-aortique dont le diamètre est donné ainsi que la longueur.
- la pression aortique moyenne donnée par les résultats.
- les informations qui concernent l'anesthésie.

Nous utilisons les noms de paragraphes et de composantes disponibles c'est-à-dire ceux qui servent déjà pour la description des documents.

P	C	V
PATT	SPEC AGE WGHT	DOG <3 <10:20>
	SPEC	CALF
BALPG	DIAM LENG	
RESUL	MAOP	
ANEST		

Le lecteur reconnaîtra la structure que nous avons donnée à la partie liste de paragraphes des documents, avec une différence de présentation. Les vecteurs sont marqués par des traits horizontaux (qui remplacent les séparateurs # et / du langage de description). Les traits verticaux qui définissent les colonnes permettent de remplacer quant à eux les séparateurs = et ;. Un début de paragraphe est signalé par la présence de son nom en colonne P et une composante occupe une seule ligne.

Dans l'exemple, le paragraphe PATT a deux vecteurs. Le premier a trois composantes (SPEC, AGE, WGHT), le deuxième une seule (SPEC).

On remarquera qu'une liste de valeurs (colonne V) peut être remplacée par des blancs (on veut faire une sélection sur un nom de composante seulement).

D'autre part, un paragraphe n'a pas nécessairement une liste de vecteurs comme le paragraphe ANEST de l'exemple le montre. La sélection porte alors uniquement sur son nom.

Il existe 8 configurations différentes pour une ligne suivant que ses parties P, C et V soient blanches ou pas. Le tableau ci-dessous les passe en revue. Un "0" représente une partie blanche, un "1" une partie non blanche.

P	C	V	
0	0	0	correspond à un trait horizontal par convention
0	0	1	combinaison interdite (erreur)
0	1	0	composante sans liste de valeurs (sélection d'un <u>nom</u> de composante)
0	1	1	composante avec liste de valeurs
1	0	0	paragraphe sans vecteurs (sélection d'un <u>nom</u> de paragraphe)
1	0	1	combinaison interdite (erreur)
1	1	0	première composante d'un paragraphe sans liste de valeurs
1	1	1	première composante d'un paragraphe avec liste de valeurs

Les règles d'écriture d'une suite de lignes se résument à : (avec des notations évidentes)

1. Une ligne de type lxy est obligatoirement précédée d'un trait horizontal (ligne 000)
2. Une ligne de type lxy ne peut pas être suivie d'une autre ligne de type lxy.
3. Une ligne de type 100 est obligatoirement suivie d'un trait horizontal.
4. Un trait horizontal ne peut pas être suivi d'un autre trait horizontal.

Nous retrouvons également une autre règle analogue à celle donnée pour la description des documents :

Les listes de valeurs des composantes d'un même vecteur doivent avoir le même nombre de valeurs.

Exemple.

P1	A	1, 2, 3
	B	5, 6, 7
	C	
	D	8, 9, 10

où P1 est un nom de paragraphe et A, B, C, D des noms de composantes (la composante C n'a pas de liste de valeurs et répond à la dernière règle énoncée).

A - 2. Vecteur pertinent.

Etant donné un vecteur de la question, il nous faut préciser quels sont les vecteurs des documents qui y répondent. Un tel vecteur est qualifié de "pertinent".

Le vecteur VD d'un document est pertinent pour le vecteur VQ de la question si les trois conditions suivantes sont satisfaites:

1. Le nom de paragraphe auquel appartient VQ existe dans le document
2. Pour chaque composante de VQ :
 - a) son nom existe dans VD
 - b) Si elle possède une liste de valeurs alors chaque valeur de cette liste est en accord avec une valeur de la composante de même nom de VD sinon le point b est toujours satisfait.
3. Les valeurs de même rang de VQ sont en accord avec des valeurs de même rang de VD, ces rangs n'étant d'ailleurs pas nécessairement égaux (Rappelons que le rang d'une valeur est sa place dans la liste de valeurs dans laquelle elle figure).

Nous préciserons le sens de "être en accord" plus loin. Nous allons illustrer d'abord ces conditions de pertinence d'un vecteur par des exemples simples.

Exemple. Vecteur de la question :

P1	A	1, 2
	B	10, 20
	C	

Vecteurs du paragraphe P1 du document :

A = 3, 2, 4, 1 ;

X = 11, 12, 13, 14 ;

C = 21, 22, 23, 24 ;

B = 30, 20, 40, 10

vecteur pertinent

A = 1, 2 [#] ;

B = 10, 20 ;

X = 40, 50

#

vecteur non pertinent

(condition 2 - a non satisfaite)

B = 10, 20, 40 ;

A = 3, 2, 1 ;

C = 21, 22, 23

#

vecteur non pertinent
(condition 3 non satisfaite)

Remarque.

Nous attirons l'attention des utilisateurs sur le fait que lorsque les listes de valeurs de la question deviennent longues, le système d'interrogation devient très sélectif. Il est primordial de rédiger correctement ses questions. Prenons la question suivante :

PATT	SPEC WGHT	DOG, CALF
------	--------------	-----------

Pour qu'un vecteur du document soit pertinent, il faut entre autres conditions que les mots DOG et CALF apparaissent dans la même liste de valeurs de la composante SPEC, ce qui est extrêmement sélectif. En particulier si dans le même document un des vecteurs parle de chiens (et de leur poids) et un autre de veaux (et de leur poids), aucun des deux n'est pertinent pour le vecteur de la question. Par contre les deux vecteurs de la question suivante auraient chacun une réponse positive :

PATT	SPEC WGHT	DOG
	SPEC WGHT	CALF

Remarque.

Pour un paragraphe de la question qui ne possède pas de liste de vecteurs, il suffit que son nom existe dans le document pour que la sélection soit pertinente.

Il nous reste à définir ce que sont des valeurs "en accord".

A - 3. Valeurs "en accord".

La définition de "en accord" est indispensable pour pouvoir vérifier la condition b - 2 de pertinence d'un vecteur d'un document pour un vecteur d'une question.

A la réflexion, il nous est apparu que l'adjectivation des valeurs d'une question n'était pas souhaitable pour deux raisons :

- dans de nombreux cas, elle ne présente aucun intérêt. Par exemple : sélectionner une liste de deux valeurs numériques dont la différence n'est pas significative (indication faite par adjectivation au moment de la description des documents).

- dans d'autres cas, l'adjectivation conduit à trop de sélectivité. Une demande telle que MYOC-INFARCT (ACUTE) exigerait que l'adjectif ACUTE soit présent dans le document, ce qui est très sélectif.

Nous avons donc interdit, tout au moins pour le moment, l'adjectivation des valeurs d'une question.

Une conséquence est l'élimination des valeurs standard normalement utilisées en tant qu'adjectifs (P, NS). Quant aux valeurs standard E, I, A, D, PC, elles ne présentent aucun intérêt dans une demande. Ceci est clair si l'on se rappelle que l'on entrainerait la recherche de ces valeurs dans les documents qui, rappelons-le, existent car on n'a pas pu faire de description plus précise.

L'adjectivation et les valeurs standard de la description des documents ne sont évidemment pas inutiles pour autant (il en est tenu compte pour faire les traitements tels que calculs ou éditions).

Ces restrictions étant apportées, nous sommes amenés à comparer entre elles des valeurs élémentaires (c'est-à-dire non adjectivées).

Deux cas sont à envisager :

1. La valeur élémentaire de la question est un mot.

- Seuls des mots du document sont susceptibles d'être en accord avec le mot de la question à l'exclusion de tout autre type de valeur.

- Les mots du document en accord avec celui de la question sont soit le même mot, soit les descendants (du point de vue de l'arborescence de hiérarchie du lexique) de ce mot.

Comme nous l'avions signalé au moment de la description du lexique des mots, nous offrons ainsi des possibilités de sélection plus performantes. Par exemple les mots du document en accord avec le mot CARNIVORE d'une question sont

CARNIVORE

LION

TIGRE

CHIEN

en considérant l'arborescence de hiérarchie donnée page 3.20.

2. La valeur élémentaire de la question est un nombre ou une fourchette.

Les seules valeurs du document susceptibles d'être en accord avec celles de la question sont les nombres et les fourchettes. Ainsi nous limitons le tableau explicatif ci-dessous à ces deux types de valeurs. Les entrées verticales correspondent aux valeurs de la question, les entrées horizontales à celles du document. A l'intersection d'une ligne et d'une colonne nous indiquons sous quelles conditions les valeurs correspondantes sont en accord.

- nq désigne un nombre de la question.
- nd désigne un nombre du document.

	nq	<nq	>nq	<nq ₁ : nq ₂ >
nd	nd = nq	nd ≤ nq	nd > nq	nq ₁ ≤ nd ≤ nq ₂
<nd	nq ≤ nd	nd ≤ nq	nd > nq	nq ₁ ≤ nd ≤ nq ₂
>nd	nq ≥ nd	nd ≤ nq	nd ≥ nq	nq ₁ ≤ nd ≤ nq ₂
<nd ₁ : nd ₂ >	nd ₁ ≤ nq ≤ nd ₂	nd ₁ ≤ nq ≤ nd ₂	nd ₁ ≤ nq ≤ nd ₂	nd ₁ ≤ nq ₂ et nd ₂ > nq ₁

Nous justifions le choix de ces conditions par le fait que la fourchette <3 désigne une valeur "un peu inférieure à 3" et non pas l'intervalle) - ∞, 3). Si la fourchette <3 figure dans une question, la valeur <4 d'un document n'est pas en accord avec elle.

Si ces conditions ne donnent pas satisfaction à l'usage, c'est-à-dire que si l'on estime qu'elles provoquent trop de silence, on pourra les modifier en considérant que les fourchettes à un seul terme sont des intervalles dont une des bornes est infinie. Les conditions pour que deux valeurs soient en accord se traduisant alors par une intersection non vide de ces intervalles. Le système devient alors moins sélectif et provoquerait par contre du bruit. Les fourchettes <nq et <nd seraient alors toujours en accord.

Notons qu'une telle modification ne poserait aucun problème du point de vue du programme de sélection, cette éventualité ayant été prévue.

Nous pouvons constater une fois de plus qu'un tel système ne peut être mis au point sans une phase de rodage.

B - Subdivision "ordres de sélection".

Cette subdivision permet de lancer l'exécution des sélections des informations décrites dans la subdivision "texte de la question" que nous venons d'étudier.

Les ordres de sélection sont donnés dans des colonnes baptisées "colonnes de sélection". Ces colonnes sont découpées en autant de cases qu'il existe de vecteurs dans le texte de la question (ces cases ont une ou plusieurs lignes suivant le nombre de composantes du vecteur correspondant).

Donnons d'abord un exemple avant de passer aux explications :

P	C	V	S1	S2	S3	S4
PATT	SPEC	DOG	X	X	N	
	AGE	< 3				
	WGHT	< 10:20 >				
	SPEC	CALF				N
BALPG			X	N	X	X
RESUL	MAOP		1	1	X	X

B - 1. Case marquée.

a) Une case d'une colonne de sélection est dite marquée quand une au moins de ses lignes contient :

- soit une croix (X)
- soit la lettre "N"
- soit un numéro.

Une case marquée donne l'ordre d'exécuter la sélection du vecteur correspondant (ou du nom de paragraphe correspondant) de la subdivision "texte de la question"(quelle que soit la manière dont elle est marquée).

b) Un numéro j écrit sur une ligne d'une case en regard d'une composante (avec ou sans liste de valeurs) commande en outre le stockage des valeurs de la composante sélectionnée correspondante dans une zone identifiée par le numéro j. On pourra effectuer des traitements sur une telle zone par la suite (dans l'exemple, les valeurs de MAOP sont stockées dans la zone n° 1).

Le même numéro ne peut apparaître qu'une seule fois dans une même colonne de sélection.

Dans une même case, il peut y avoir plusieurs numéros (stockage des valeurs de plusieurs composantes).

Les numéros sont choisis au gré de l'utilisateur dans les limites 1 à 99 (bornes incluses).

c) La croix (X) écrite dans une case est utilisée quand on ne veut pas faire de stockage de valeurs. Il est indifférent d'écrire une ou plusieurs croix dans une case. Une croix peut être écrite dans une case qui contient déjà des numéros (elle est alors inutile, mais nous avons voulu que le marquage des cases soit très souple).

d) Dans chaque colonne de sélection, il doit exister au moins une case marquée par X ou / et des numéros.

e) La lettre N figure obligatoirement seule dans une case. Nous verrons son utilisation plus loin.

B - 2. Sélection de vecteurs d'une même expérience.

Pour demander la sélection de plusieurs vecteurs d'une même expérience (au sens défini dans le chapitre II), on marquera les cases (qui correspondent à ces vecteurs) d'une même colonne.

Dans l'exemple donné ci-dessus, dans la colonne de sélection n° 1 (S1), on veut sélectionner des chiens d'âge inférieur à 3 ans et de poids compris entre 10 et 20 kg et qui ont été traités à l'aide d'un pompage par ballonnet intra-aortique (BALPG). En outre, on veut que la pression aortique moyenne soit donnée pour ces chiens là dans les résultats. Tous les vecteurs correspondants doivent appartenir à la même expérience du document.

Pour sélectionner un vecteur indépendamment de son appartenance à une expérience, on ne marquera qu'une seule case d'une même colonne.

B - 3. Colonne de sélection pertinente.

Une colonne de sélection est dite pertinente si les conditions suivantes sont toutes satisfaites :

- a) Pour chaque case de la colonne marquée par X ou/et des numéros :
 - si la sélection porte sur un vecteur alors il existe un vecteur pertinent dans le document.
 - si la sélection porte sur un nom de paragraphe alors ce nom existe dans le document.

- b) Les résultats des sélections marquées par X ou/et des numéros appartiennent tous à la même expérience numéro p du document :
 - si un résultat est un vecteur, il appartient à l'expérience n° p.
 - si un résultat est un paragraphe, un au moins de ses vecteurs appartient à l'expérience n° p.

- c) Pour chaque case marquée par N :
 - si la sélection porte sur un vecteur alors aucun vecteur pertinent n'appartient à l'expérience n° p.
 - si la sélection porte sur un nom de paragraphe, alors :
 - soit ce nom n'existe pas
 - soit ce nom existe et aucun de ses vecteurs n'appartient à l'expérience n° p.

Remarques.

- Rappelons qu'un vecteur (d'un document) qui ne contient pas la composante EXP dans sa description concerne toutes les expériences présentes dans le document.

- Le marquage d'une case par N doit être utilisé prudemment et en toute connaissance de cause. L'utilisation de N revient à éliminer les expériences qui contiennent certains vecteurs. La colonne de sélection n° 2 de l'exemple montre une bonne utilisation de N. En clair, on veut sélectionner les chiens pour lesquels MAOP est donnée mais qui n'ont pas été traités par ballonnet intra-aortique (contrairement à la demande de la colonne n° 1).

La colonne n° 3 donne un exemple où l'utilisation de N est plus délicate. En fait, il faut bien voir quelles sont les expériences qui sont éliminées des réponses aux questions par ce marquage. Conformément à ce que nous avons dit, ce sont celles qui contiennent un vecteur pertinent pour celui de la question. Dans le cas qui nous préoccupe, le vecteur suivant du document :

SPEC = DOG ;

AGE = 8 ;

HEIGHT = <10:20>

#

ne constitue pas un critère d'élimination puisqu'il n'est pas pertinent pour celui de la question (à cause de la composante AGE).

Dans la colonne S4, on veut éliminer toutes les expériences qui parlent de veaux (bonne utilisation de N).

Nous demandons aux utilisateurs d'être prudents et de n'utiliser le marquage d'une case par N qu'à bon escient. Les deux cas les plus intéressants étant le marquage soit d'un nom de paragraphe, soit d'un vecteur à une seule composante.

Remarquons enfin que N ne peut pas figurer seul dans une colonne ce qui n'aurait pas de sens (ceci est exprimé par 1 - d).

B- 4. Réponses à une colonne de sélection.

Nous appelons réponse à une colonne de sélection, un ensemble de vecteurs pertinents pour cette colonne (compte-tenu bien entendu des conditions de B-3).

Au moment de l'exécution des ordres de sélection d'une colonne, toutes les réponses sont recherchées.

Remarques.

La composante de nom EXP ne doit pas être utilisée pour rédiger les questions. En effet, rechercher les vecteurs pour lesquels $EXP = n$ n'a pas de sens puisque le choix des numéros d'expériences est arbitraire au moment de la description des documents. Voir à ce propos le point B-2 ci-dessus.

Pour un même vecteur de la question, il peut exister plusieurs vecteurs pertinents du document (puisque toutes les réponses sont recherchées). Par conséquent, il peut exister plusieurs composantes pertinentes pour une même composante de la question. Si un stockage des valeurs a été demandé toutes le sont dans la zone désignée (qui est découpée en sous-zones contenant chacune les valeurs d'une liste, mais ceci est transparent à l'utilisateur).

C - Subdivisions "traitements" et "ordres de traitements".

Les traitements à entreprendre sont décrits dans la partie "traitements" de la table. Il sont séparés les uns des autres par des traits horizontaux.

Les demandes d'exécution des traitements se font dans la subdivision "ordres de traitements". Celle-ci est découpée en autant de colonnes qu'il y a de colonnes de sélection (par le prolongement de celles-ci). Une colonne de traitement est elle-même découpée en autant de cases qu'il y a de traitements prévus dans la subdivision correspondante.

Exemple.

texte de la question	X	X	
		X	
	X		X
			X
traitement n°1		2	X
traitement n°2	X	1	
traitement n° 3			X

Une colonne de traitement peut avoir une ou plusieurs cases marquées par des croix. L'ordre d'exécution des traitements correspondants se fait alors de haut en bas (pour la troisième colonne, il y a exécution du traitement n° 1, puis du traitement n° 3). Pour obtenir un ordre différent, on marque les cases à l'aide de numéros croissants (à partir de 1) (pour la colonne n° 2, il y a exécution du traitement n° 2 puis du traitement n° 1).

Nous parlerons des traitements eux-mêmes un peu plus loin (en E).

D - Fonctionnement de la table de sélection et de traitements.

Nous allons voir ici comment se succèdent les exécutions des sélections et des traitements.

1. Les colonnes de sélection sont explorées successivement de gauche à droite sauf contre ordre (donné éventuellement dans un traitement) et jusqu'à la dernière.
2. Si une colonne de sélection est pertinente, les traitements marqués correspondants sont exécutés. Puis la règle 1 est appliquée.
3. Si une colonne de sélection n'est pas pertinente, la règle 1 est appliquée (sans exécution des traitements marqués).

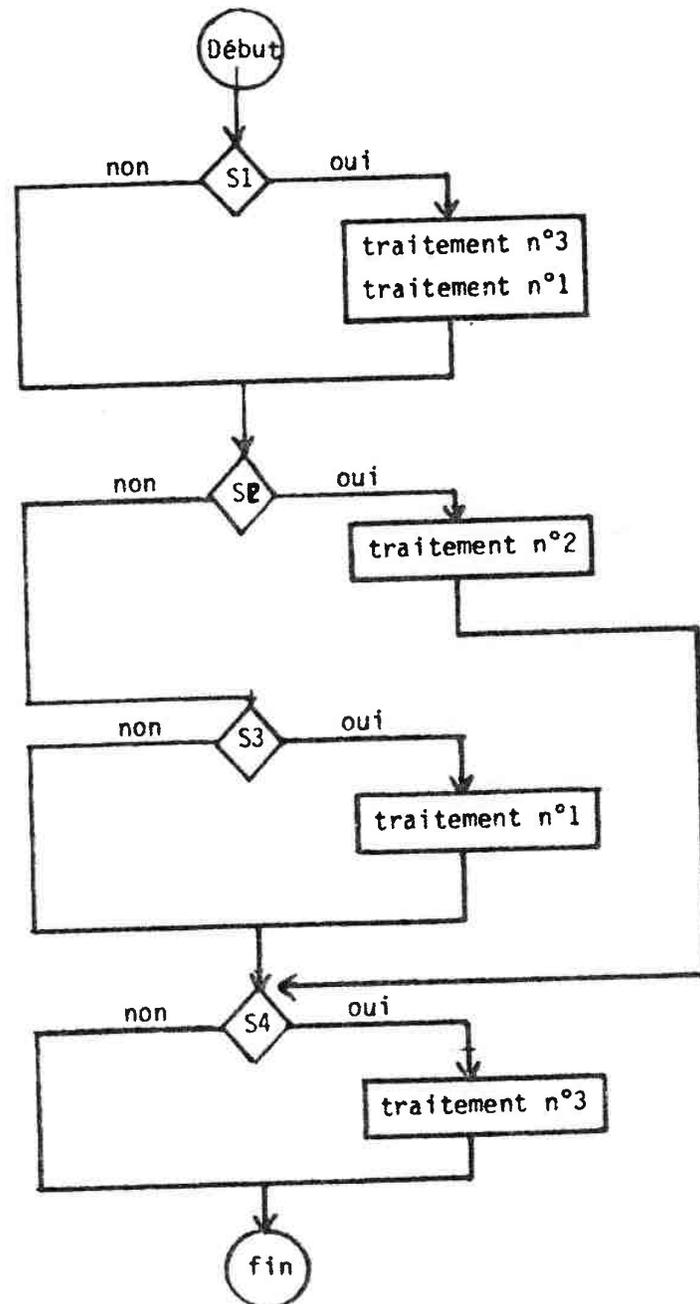
Soit la table suivante :

texte de la question	S1	S2	S3	S4
	ordres de sélection			
traitement n° 1	2		X	
traitement n° 2		X		
traitement n° 3	1			X

Imaginons que dans le traitement n° 2, il y ait une instruction d'aller à la colonne de sélection n° 4. Le déroulement des opérations se ferait alors comme le montre l'organigramme de la page 4.17.

Organigramme du déroulement de la question de la page 4.16.

La notation  signifie : la colonne de sélection n°i est-elle pertinente ?



E - Les traitements.

Nous parlons ici des traitements qui figurent dans la table de sélection et de traitements et que nous avons appelés locaux car ils sont exécutés au niveau d'un document. (voir le fonctionnement de la table en D).

Comme nous l'avons déjà signalé, cette partie "traitements" n'est pas entièrement définie à ce jour et nous donnerons ici les grandes lignes qui guideront la définition de ces traitements.

Les traitements sont des listes d'instructions dont nous allons spécifier rapidement les différents types.

E - 1. Instructions de gestion de la table de sélection.

Les deux instructions définies ci-dessous permettent de modifier l'ordre normal d'exploration des colonnes de sélection.

a) Instruction STOP.

Lorsque cette instruction est exécutée, elle entraîne l'abandon de l'étude du document en cours (toutes les colonnes de sélection ne sont donc pas nécessairement explorées).

b) Instruction SELECT (i).

L'exécution de cette instruction provoque l'exploration de la colonne de sélection n° i (les colonnes sont numérotées implicitement à partir de 1 et de gauche à droite). Le numéro i doit être strictement supérieur au numéro de la colonne de sélection qui a entraîné l'exécution du traitement dans lequel cette instruction se trouve. Dans l'exemple donné en D page 4-15, le traitement n°2 contenait : SELECT (4).

E - 2. Instruction de répétition d'une liste d'instructions.

Rappelons qu'une sélection peut fournir plusieurs réponses pour une même colonne. Nous avons introduit en conséquence une instruction de répétition d'un même traitement pour chacune des réponses obtenues.

Cette instruction (PCR) aura la syntaxe suivante :

PCR (liste d'instructions autre que PCR)

L'effet de cette instruction est d'exécuter la liste d'instructions entre parenthèses pour chaque réponse.

E - 3. Autres instructions.

Nous pensons définir les instructions suivantes :

1. Instructions d'édition.

Ces instructions devront permettre d'éditer des paragraphes, des composantes, des zones de stockage de valeurs, des résultats de calculs, des libellés, etc...

2. Instructions de conservation de valeurs locales.

Une valeur locale est :

- soit le résultat d'un calcul d'un traitement local,
- soit une valeur stockée au moment de la sélection (case marquée par un numéro) et dont la durée de vie n'excède pas celle de l'étude d'un document.

Pour pouvoir faire des traitements sur des valeurs provenant de l'ensemble des documents, nous introduirons des instructions qui permettent de conserver les valeurs locales dans des zones que nous appelons globales. Ces zones globales sont donc "remplies" au fur et à mesure du déroulement du fichier des documents et constituent des tableaux de résultats (exploitables par un traitement global).

On veut relever sur l'ensemble des documents les valeurs de MAOP obtenues sur des chiens traités par ballonnet intra-aortique de capacité 20 cc afin d'en établir la moyenne.

PATT	SPEC	DOG	X
BALPG	BALCA	20	X
RESUL	MAOP		1
PCR (conserver la zone 1 dans la zone globale numéro 1)			X

Traitement global.

- faire la moyenne des valeurs de la zone globale n° 1.
- Editer cette moyenne

3. Instructions de traitements.

Ces instructions permettent de faire des calculs sur des listes de valeurs stockées dans des zones locales (Par exemple : calculs de moyennes, recherches de maximums et de minimums).

4. Instruction de style "langage évolué".

S'il s'avère que les traitements demandés par les utilisateurs deviennent complexes et variés, les instructions précédentes ne suffisent plus à résoudre leurs problèmes. Dans ce cas nous introduirons des instructions de style FORTRAN ou COBOL par exemple (avec affectation, calcul arithmétique, tests, bouclages, etc...).

Autant que possible, nous aimerions éviter l'introduction de ces instructions car nous avons le souci de fournir un outil de travail adapté aux compétences de l'utilisateur.

Exemple.

On veut connaître la pression aortique moyenne (MAOP) minimum ayant été relevée sur des chiens de plus de 10 kg ayant eu un infarctus du myocarde dans deux cas de traitements : avec ou sans pompage par ballonnet intra-aortique.

PATT	SPEC	DOG	X	X
	WGHT	10		
PROC	COND	MYOC-INFARCT	X	X
BALPG			X	N
RESUL	MAOP		1	1
PCR (rechercher et éditer la valeur minimum de la zone n° 1)			2	2
Editer le texte 'AVEC BALLONNET'			1	
Editer le texte 'SANS BALLONNET'				1

Pour la colonne de sélection n° 1, on demande d'imprimer d'abord le texte 'AVEC BALLONNET' puis, pour chaque réponse d'imprimer la valeur minimum de MAOP. Pour la deuxième colonne, on veut imprimer le texte 'SANS BALLONNET', puis faire le même traitement.

IV. 3 - ENTREE D'UN TRAIN DE QUESTIONS EN MACHINE.

Nous venons d'étudier la partie essentielle d'une question qui est la table de sélection et de traitements. Comme nous l'avions annoncé dans l'introduction, une question comporte en outre les parties suivantes :

- 1 - Identification de la question
- 2 - Limites d'exploitation du fichier des documents
- 3 - Sélection sur l'entête d'un document.

Les parties 2 et 3 ne sont d'ailleurs pas nécessairement présentes.

Nous allons donner en même temps que des explications, la manière de présenter les questions sur un bordereau de perforation (les questions sont entrées par cartes).

A - Identification de la question.

La première ligne d'une question est toujours réservée à cet effet. L'utilisateur en dispose pour y inscrire ses noms et prénoms ou tout autre renseignement qu'il juge utile de donner (pour sa gouverne personnelle).

Cette ligne a le format suivant :

1	3	8
0 0	QUES =	chaîne de caractères

La chaîne de caractères ne peut pas se poursuivre sur une deuxième ligne.

B - Limites d'exploitation du fichier des documents.

L'utilisateur peut demander que le fichier des documents ne soit pas déroulé entièrement pour répondre à sa question. Il peut, s'il le désire, donner les limites d'exploitation suivantes :

1. Le nombre maximum (MAXI) de documents à étudier. L'interrogation du fichier des documents s'arrête alors (pour sa question) quand ce nombre est atteint.

2. Le nombre de documents pertinents (PERT) à partir duquel il estime être suffisamment renseigné. L'interrogation s'arrête quand ce nombre est atteint.

Ces clauses sont optionnelles. Par défaut tout le fichier des documents est étudié. Si les deux limites sont présentes à la fois, l'interrogation s'arrête dès que l'une d'elles est atteinte.

Règles de présentation.

- Ecrire un seul de ces renseignements par ligne
- Numérotage des lignes dans les colonnes 1 et 2 (à partir de 01 car la première carte de la question porte le numéro zéro. Se reporter à l'identification de la question).
- Le nombre maximum de documents à étudier est donné par :
MAXI = nombre entier sans signe
- Le nombre satisfaisant de documents pertinents est donné par :
PERT = nombre entier sans signe
- Ces limites sont données à partir de la colonne 3
- Si les deux limites sont présentes, leur ordre d'écriture est indifférent.

Exemple.

1	3	8
0 0	QUES =	SZTUKA - ETUDE DE BALPG -
0 1	MAXI =	300
0 2	PERT =	50

C - Sélection sur l'entête.

Parmi les renseignements donnés dans l'entête d'un document, nous avons retenu la référence et l'année de parution comme éléments de sélection.

C - 1. Sélection sur la référence.

Rappelons que la référence est un numéro à trois chiffres qui désigne le nom d'un livre ou d'une revue.

L'utilisateur peut demander que l'on ne tienne compte que des documents ayant paru sous certaines références dont il précise les numéros sur la

première ligne vierge du bordereau (à partir de la colonne 3) de la manière suivante :

REFS = liste de références.

Les différentes références de la liste sont séparées entre elles par des virgules.

Une telle ligne ne peut pas se poursuivre sur la ligne suivante ce qui permet donc de donner au maximum une liste de 18 références (nombre largement suffisant). Il y a numérotage d'une telle ligne.

Exemple.

1	3	8
0 0	QUES =	DUPONT
0 1	REFS =	101, 104, 105

C - 2. Sélection sur l'année de parution.

On peut demander que l'interrogation ne se fasse que sur les documents dont la parution est antérieure, postérieure ou égale à une certaine année.

On écrira sur la première ligne libre du bordereau (à partir de la colonne 3) :

- soit PARU = année
- soit PARU * année
- soit PARU < année

où "année" est un nombre entier sans signe de 4 chiffres.

Exemple.

1	3	8
0 0	QUES =	DURAND
0 1	MAXI =	200
0 2	PARU >	1970

Seuls sont pris en compte les documents dont la parution est postérieure à 1970.

Remarque.

Si les deux éléments de sélection REFS et PARU sont simultanément présents, l'ordre dans lequel ils sont écrits, est indifférent.

D. - Présentation de la table de sélection et de traitement sur des bordereaux.

Nous avons pensé qu'il serait agréable pour l'utilisateur de présenter cette table sur des bordereaux de manière analogue à la représentation que nous en avons faite en IV. 2. En contre partie, cette façon de procéder impose des limites à la taille de cette table (à cause du format du bordereau). En particulier, le nombre maximum de colonnes de sélection est 10 (ce qui semble largement suffisant par ailleurs).

- Les deux premières colonnes servent au numérotage des lignes.
- Les colonnes 3 à 7 sont réservées pour les noms de paragraphes (zone P).
- Les colonnes 9 à 13 sont réservées par les noms de composantes (zone C).
- Les colonnes 15 à 60 sont réservées pour les listes de valeurs (zone V) qui ne peuvent pas se prolonger sur une deuxième ligne.
- Les colonnes 61 à 80 sont réservées pour dix colonnes de sélection (chacune ayant donc 2 caractères de largeur).
- Les traits horizontaux de la table sont remplacés par une ligne blanche (le numérotage étant toutefois respecté) ce qui explique notre convention sur les types des lignes (voir A - 1, page 4.3).
- Les colonnes 8 et 14 restent blanches et matérialisent les traits verticaux qui délimitent les zones P, C et V.
- La subdivision "traitement" de la table est annoncée par les caractères "TRAIT" écrits dans les colonnes 3 à 7 après le trait horizontal (ligne blanche) qui termine la subdivision "texte de la question".
- Nous n'avons pas encore précisé la façon d'entrer les traitements puisqu'ils ne sont pas entièrement définis.

1	3	8	9	13	15	61	80
↑	↑		↑				
numérotage des lignes		zone P		zone C		zone V (listes de valeurs)	
						10 colonnes de sélection	
		TRAIT					

Remarque.

On peut envisager, pour l'avenir, de faire imprimer des bordereaux spécialement étudiés pour entrer des tables de sélection et de traitements en machine.

E - Train de questions.

Comme nous l'avons signalé dans l'introduction, l'interrogation du fichier des documents se fait en batch, c'est-à-dire qu'un train de questions est rassemblé et traité en une seule fois.

Après perforation sur cartes des questions décrites sur les bordereaux (suivant les directives que nous venons de donner), celles-ci sont rassemblées en un fichier (TRAIN) qui est pris en compte par le système d'interrogation. Ce fichier peut contenir un nombre quelconque de questions.

Nous allons dire quelques mots sur la réalisation de ce système d'interrogation.

IV. 4 - REALISATION DU SYSTEME D'INTERROGATION.

Actuellement, la réalisation du système d'interrogation est faite en partie seulement puisque les traitements que nous voulons introduire dans les questions ne sont pas entièrement définis à ce jour.

Nous avons toutefois voulu démarrer rapidement l'exploitation du fichier des documents en nous limitant à la partie "sélection" des questions. Ainsi donc le système peut être utilisé comme un système strictement documentaire, c'est-à-dire qui fournit une liste de documents pertinents pour des critères formulés dans des questions.

Un train de question est exploité en deux phases :

1. Création d'un fichier de questions codées (FQST)
2. Exploitation du fichier des documents à partir des questions codées (de FQST).

Nous donnons l'organigramme général du système d'interrogation, page 4.28

A - Création d'un fichier de questions codées (FQST).

Les questions d'un train sont codées et conservées dans le fichier FQST en vue de la phase d'interrogation proprement dite. Ce travail est effectué par un programme de nom QST. Dans cette phase :

- Il y a vérification de la syntaxe (présentation correcte de la table de sélection et de traitement). En cas d'erreurs des messages sont édités.

- Le codage des questions est fait par consultation du lexique des noms de paragraphes et de composantes, du lexique des mots et également du fichier qui rend compte de la structure hiérarchique des mots (voir III. B - 3). Les descendants d'un mot sont adjoints à celui qui existe dans la question qui se trouve donc complétée. Il en sera donc tenu compte dans la phase de sélection.

- Chaque question correcte est conservée sous sa forme codée dans le fichier FQST et en constitue un enregistrement logique.

- Nous ne décrivons pas les enregistrements de FQST ici. Disons simplement qu'ils sont organisés par colonne de sélection puisque l'exploration de la table de sélection se fait colonne par colonne (voir fonctionnement de la table en IV. 2).

- Le programme QST produit un listing des questions avec éventuellement des messages d'erreurs. Nous avons soigné ces éditions et en particulier les tables de sélection sont présentées avec des "traits".

B - Exploitation du fichier des documents à partir de questions codées.

Cette exploitation des documents est faite par un programme de nom REP qui répond à chacune des questions/^{de}FQST. Le programme fait les sélections demandées par les questions.

Pour chaque question :

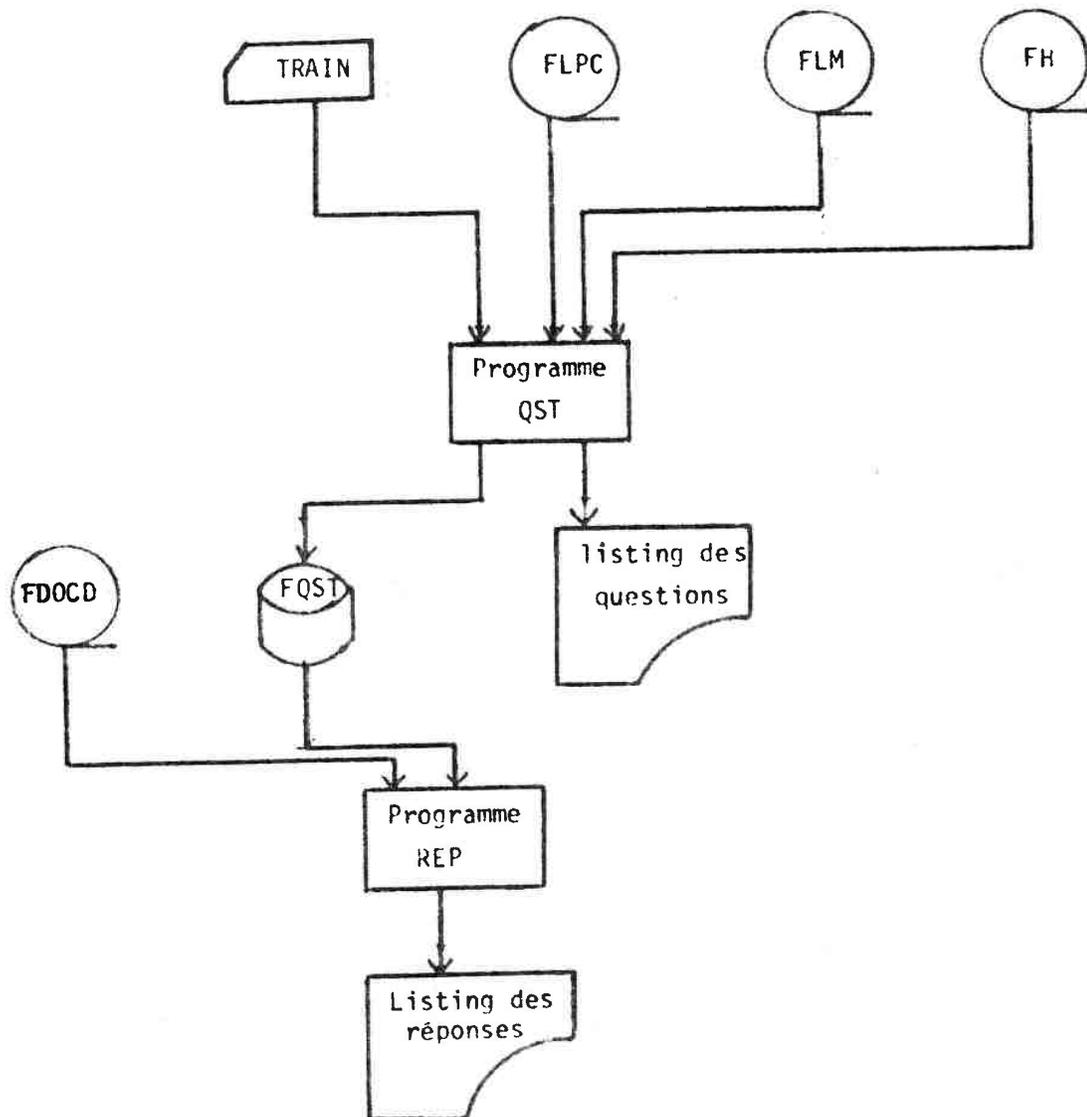
- Il y a édition de toute la partie identification des documents pertinents.

- Pour chaque document pertinent il y a édition des numéros des colonnes de sélections pertinentes.

- Il y a édition du nombre total des documents pertinents et du nombre total de documents étudiés.

C - Exemple.

Le lecteur trouvera un exemple simple de question et la réponse obtenue dans l'état actuel du système d'interrogation à la page 4.29.

Organigramme général du système d'interrogation.

IDENTIFICATION DE LA QUESTION : QUES= EXEMPLE

** DOCUMENT PERTINENT **
TITRE : DIRECT AND REFLEX VASCULAR EFFECTS OF IN
TRA AORTIC BALLOON COUNTERPULSATION
IN DOGS
INDICATIF : 1001
REFERENCE : 001
ANNEE : 1971
TOME : 221
PAGE DEBUT : 0748
PAGE FIN : 0753

COLONNES DE SELECTION PERTINENTES
01
02

* REPONSE TERMINEE * DOCUMENTS PERTINENTS : 001 DOCUMENTS ETUDIES : 004

*** FIN DE TRAVAIL ***

CONCLUSION

Qu'il nous soit permis, à l'issue de ce travail, d'attirer l'attention du lecteur sur les quelques points suivants :

- Dans son état actuel notre système de documentation n'est pas pleinement opérationnel et sa mise au point se poursuit.
 - La réalisation d'un tel système est une oeuvre de longue haleine; elle nécessite des entretiens fréquents et une collaboration approfondie entre "l'utilisateur-médecin" et l'informaticien, afin de résoudre les problèmes qui se posent tant sur le plan de la communication que sur celui de la définition d'objectifs précis.
 - Il semble qu'on gagnerait à travailler en équipe plus nombreuse.
 - Sur la base d'une collaboration suivie au sein d'une telle équipe on pourra alors parfaire ce système et le compléter.
 - Bien que ce système ne prétende pas être universel, nous espérons toutefois qu'il pourra être appliqué, sans gros problèmes, à d'autres domaines que celui de l'Assistance Cardio-Circulatoire.
-

BIBLIOGRAPHIE

- 1 - K.F. BONDEN, I.R. MACCALUM
Toward a self-generating medical dictionary
Journées d'Informatique Médicale, Toulouse, 1972, p. 69.
- 2 - J. CHAUMIER
Systèmes informatiques de Documentation.
Entreprise Moderne d'Edition, Paris 1972.
- 3 - C. DE BARY
Thèse de Doctorat de Spécialité relative au traitement
de fiches de relevés géologiques, Nancy (à paraître).
- 4 - J.C. GARDIN, R.C. CROS, F. LEVY
L'automatisation des recherches documentaires.
Un modèle général "LE SYNTOL".
Gauthiers Villars, Paris, 1968.
- 5 - P. GERMAIN
Système de gestion et d'exploitation documentaire d'un
corpus de dossiers médicaux.
Thèse de Doctorat de Spécialité, Nancy, 1972.
- 6 - R. MAHL
Cours d'informatique software. Structure de données.
U.E.R. des Sciences, St-Etienne.
- 7 - C. PAIR
Cours : Structure de l'information
Théorie des langages
Théorie des graphes
Nancy 1970
- 8 - J.P. TRYSTRAM
La documentation automatique.
Paris, Dunod, 1971.
- 9 - L. FOSSIER, M. CREHANGE
Essai d'exploitation sur ordinateur des sources
diplomatiques médiévales.
Annales - Economies, Sociétés, Civilisation. N° 1, Janvier-
Février 1970, Paris, Armand-Colin.



NOM DE L'ETUDIANT : SCHNEFF Yves

NATURE DE LA THESE : Doctorat de Spécialité en Mathématiques Appliquées.

VU, APPROUVE

& PERMIS D'IMPRIMER

NANCY, le 24 Juin 1974

LE PRESIDENT DE L'UNIVERSITE DE NANCY :



J.R. HELLOY