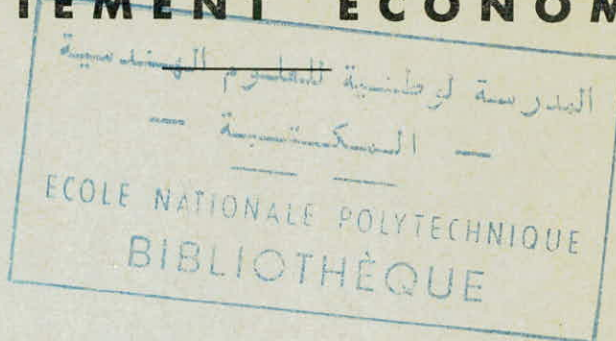


UNIVERSITÉ D'ALGER

2/74

ÉCOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT ECONOMIE



ANALYSE SEMANTIQUE DE L'INFORMATION ECONOMIQUE

THÈSE DE FIN D'ÉTUDES



SUJET PROPOSÉ PAR
M^r DOLIATOVSKI Dr Es-Sciences

ÉTUDIÉ PAR M^r
BRANCI Benabdellah
et
EL-KETROUSSI Mohamed

PROMOTION 1974

A, CUS DEDIONS

ETTE ETUDE

DES PARENTS

ET AMIS

EXCLU DU PRÊT

REMERCIEMENTS

Que toutes les personnes qui ont contribuées
à la Formation trouvent ici l'expression de nos
vifs remerciements. Nous rendons Hommage particulièrement
à /sieur V. DOLIATOVSKI , pour l'aide et les conseils
qu'il a prodigués , ainsi qu' à //)essieurs TERKOUCHE.A,
DAMI sans qui ce projet ne serait jamais réalisé.



ANALYSE SEMANTIQUE
DE
L'INFORMATION ECONOMIQUE

INTRODUCTION

CHAPITRE I: ANALYSE FORMELLE ET SYSTEME FORMEL

- 1-1: analyse logique
- 1-2: analyse syntaxique
- 1-3: analyse semantique

CHAPITRE II: DESCRIPTION DES INFORMATIONS ECONOMIQUES

- 2-1: morphologie de l'information economique
- 2-2: les éléments de l'information
- 2-3: classification-relations logiques
- 2-4: modèle logique de l'information
- 2-5: notions de thesaurus

CHAPITRE III: ANALYSE STATISTIQUE DE L'INFORMATION

- 3-1: chaines de traitement air-algerie
- 3-2: analyse des documents
- 3-3: liste des expressions
- 3-4: liste des objets et des prédicats
- 3-5: construction du thesaurus

CHAPITRE IV: SYSTEME SEMANTIQUE DE TRAITEMENT

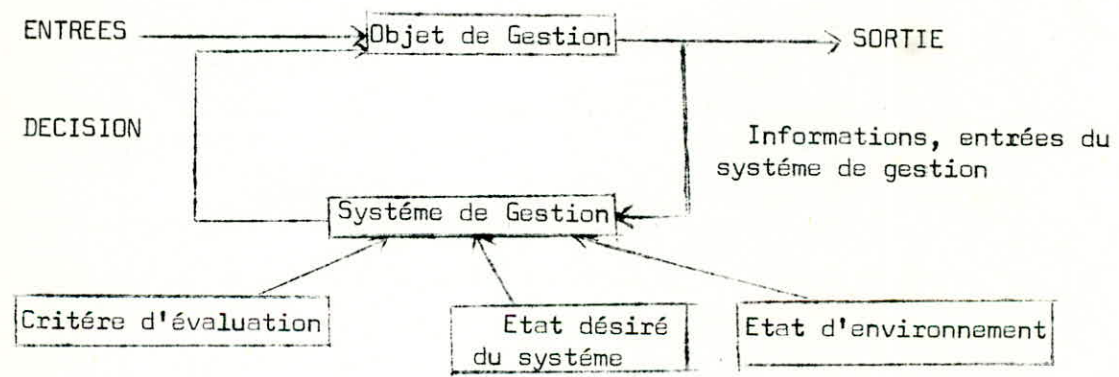
- 4-1: configuration générale du système
- 4-2: constitution des appels
- 4-3: organisation des données
- 4-4: fonctionnement du système

CHAPITRE V: EXEMPLE D'APPLICATION

CONCLUSION :

--- INTRODUCTION ---

Toute l'activité d'une entreprise se reflète par un mouvement d'informations dans le temps et l'espace. L'information circule entre des éléments de trois types : des sources G, des transformateurs L et des consommateurs C. Ces chaînes G-L-C peuvent être distinguées pour chaque traitement de l'information, le transformateur L réalise la transmission et le traitement de l'information ainsi qu'éventuellement la prise de décisions. Comme on le constate, la base d'étude et de construction d'un système rationnel de gestion pour l'entreprise est l'information, donc d'une part la réception, la circulation et le traitement de l'information est rapide ; plus le système de gestion de l'entreprise aura une grande fiabilité et une capacité d'auto-organisation ainsi que de bonnes caractéristiques dynamiques d'une part et structurales d'autre part, l'entreprise ne peut évoluer que si son système de gestion fonctionne bien .



Le Schéma ci-dessus nous résume le fonctionnement d'un système de gestion l'information y joue un grand rôle. Aussi toutes les études qui ont été entreprises pour analyser d'abord et améliorer ensuite un système de gestion tiennent compte en grande partie, de l'information qui y circule parmi celles-ci on peut citer : Méthodes matricielle du CEM Moscou, Méthode graphe-analytique de l'institut des problèmes de gestion-Moscou, Méthode analytique Séquentielle (V.Joliatovski), Méthode de Blapré (IBM) etc... Mais dans toutes ces études deux approches ont été principalement développées :

- L'étude quantitative du flux de l'information économique :

L'entreprise est alors représentée comme un réseau informationnel pour lequel on définit les échanges entre les différents éléments. Le système de gestion de l'entreprise est décomposé en sous-systèmes, chaque sous-systèmes ^à de décision qui ont des influences sur les éléments de décision des autres sous-systèmes.

Après cette décomposition on détermine, par des méthodes d'enquête, le volume de travaux pour chaque sous-système, la répartition de ce volume entre les différents éléments du sous-système, la productivité de ces éléments ainsi que les ressources humaines et techniques utilisées.

Ce sont donc des caractéristiques quantitatives du système qu'on aura après cette analyse. Ces caractéristiques nous permettent de déterminer "les goulots d'étranglement" du système et d'aboutir à une meilleure répartition des travaux entre les différents éléments de l'entreprise.

- L'étude fonctionnelle de l'information économique :

Cette étude consiste à considérer l'entreprise comme ayant une fonction globale F à exécuter. Cette fonction F est décomposée en fonctions partielles ou sous-fonctions f_i ; à chaque fonction f_i , on attache un certain nombre de tâches : Z_{mi} qu'il faut exécuter. Ces tâches peuvent avoir des points communs et des liaisons intermédiaires et sont représentées par des documents qui s'y rattachent :

$$\forall Z_{mi}, \exists e_{mi} / f_i \cap f_j = \emptyset \text{ et } Z_{mi} (e_{mi}) = S_{mi}$$

e_{mi} et S_{mi} sont des documents d'entrée et de sortie d'une tâche Z_{mi} . Les répartitions des fonctions et des tâches d'une structure de gestion doivent être organisées de telle façon à avoir des caractéristiques du système désiré. Ces deux approches nous permettent de déterminer le flux d'information nécessaire pour chaque type de décision et pour les tâches types de gestion telles que : la Comptabilité, la Gestion du personnel, etc.... et surtout d'accélérer " le diagnostic du système de gestion " et le processus de son amélioration ou de sa reconstruction.

Mais pour mieux cerner encore et résoudre les problèmes de gestion il faudrait analyser un autre côté du système de gestion : le côté sémantique de l'information. Cette autre approche est due au fait que chaque système de gestion utilise un langage : celui contenu dans les documents et que nous appellerons langage de l'information économique ou LIE. Les documents utilisant ce langage ont leur structure, leur morphologie, leur algèbre et leur procédure de traitement.

Cette troisième approche est l'objet de notre thèse de fin d'études. Dans un premier temps nous allons essayer d'analyser les documents utilisés dans le système de gestion de la compagnie Nationale AIR-Algérie et essayer d'établir un embryon théorique de l'analyse de ces documents, donc de simplifier, de rationaliser et d'effacer les répétitions des informations, d'établir et de formaliser le langage utilisé dans les documents. Dans un deuxième temps, nous appliquerons les résultats trouvés pour un système de traitement sémantique de l'information économique.

ANALYSE FORMELLE ET SYSTEME FORMEL :

I-1- Analyse Logique :

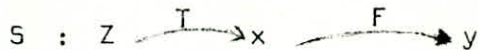
Pour tout système de gestion on distingue des paramètres d'entrée qui sont des variables extérieures au système et qui influent sur son fonctionnement et les paramètres de sortie qui sont le résultat du fonctionnement du système. Dans le cas général, on peut définir l'état d'un système de gestion à l'aide des paramètres entrés $X=(x_1, x_2, \dots, x_n)$ et chaque paramètre de sortie est lié avec X par des liaisons fonctionnelles :

$$y_i = F_i(X)$$

Mais à l'entrée du système, il existe encore m paramètres $Z=(z_1, z_2, \dots, z_m)$ qui influent sur l'objet de gestion et dont dépend X tel que :

$$x_k = T_k(Z)$$

Le système de gestion se représente alors comme une partie T qui transforme Z à X et des transformations fonctionnelles $F : x$ à y



Nous nous intéressons à la transformation : $y_i = H(Z)$ tel que $H=F \cdot T$ c'est à dire que pour analyser logiquement un système de gestion on doit définir les ensembles X, Y et Z , choisir les transformations F, T et H et définir les réflexions existantes entre ces ensembles .

Ainsi l'étude du flux de l'information d'entrée d'un système est importante pour l'analyse du système de gestion économique et c'est ainsi que comme méthode principale d'analyse on utilise les graphes et la théorie des réseaux.

Les sommets du réseau sont $X=(x_1, x_2, \dots, x_n)$ donc un ensemble de n éléments de nature différente, T sera alors une correspondance de x dans x mettant en relation chaque sommet x avec un certain nombre d'autres sommets. Les arcs du graphe $G(X, T)$ nous indiquent le flux d'informations qui circulent en temps et en liaisons avec les tâches fonctionnelles de gestion .

L'étude de ces flux d'informations nous renseigne sur la quantité d'information qui circulent entre les divisions, l'organisation logique de ces informations et sur la quantité d'information qui s'attache à chaque fonction du système de gestion de l'entreprise pour pouvoir les améliorer et choisir les moyens techniques pour le traitement de ce flux d'information avec les caractéristiques désirées (rapidité de calcul, degré d'intégration, forme de représentation etc...) Sur un réseau informationnel on pose comme contraintes principales: des tâches de minimisation de circulation des documents, la diminution de leur redondance et leur orientation vers une prise de décision rapide.

D'autre part ce réseau informationnel nous enseigne aussi sur la structure de l'entreprise ou plus exactement sur sa décomposition fonctionnelle c'est à dire à :

- Organiser les buts G_i et les critères k_i de fonctionnement des sous-systèmes fonctionnels.
- Déterminer les ensembles T_i des tâches qui sont nécessaires pour la réalisation des buts G_i précisés, l'union de toutes ces tâches fournit l'ensemble des tâches de gestion : $\bigcup_i T_i = T$
- Distinguer les sous systèmes fonctionnels F_i qui ont des tâches homogènes. Ces tâches T_i ont des liaisons à l'intérieur de F_i et des liaisons extérieur avec les autres sous-systèmes. Donc pour les analyser nous distinguerons:
 - 1er niveau : liaisons informationnelles et logiques entre les tâches T_i
 - 2° niveau : liaisons entre les sous-systèmes F_i .

Donc, l'analyse du réseau informationnel de l'entreprise est utile pour la construction d'un modèle de l'entreprise. Les tâches devront par la suite une base de construction de l'approvisionnement mathématique du système de gestion et la structure fonctionnelle des tâches nous permet de construire les fichiers communs des fonctions de gestion.

Pour chaque système de gestion on peut construire la liste des fonctions, $F=(f_1, f_2, \dots, f_n)$ qu'il faut réaliser pour la gestion, ces fonctions sont en général : service du personnel, approvisionnement, stockage, gestion financière, comptabilité, calcul économique etc...

En pratique toutes ces fonctions sont réalisées par les divisions fonctionnelles de l'entreprise : $S=(s_1, s_2, \dots, s_m)$. La répartition de ces fonctions entre les différentes divisions peut-être exprimée par la matrice.

	S_j	s_1	s_2	s_m
F_i					
f_1		1	0	0
f_2		1	1	0
.....	
f_n		0	1	1

Si nous considérons les éléments de la matrice (F/S) comme les ressources utilisées dans chaque division pour chaque fonction f_i , on aura la répartition des ressources par fonction :

		S			
		s_1	s_2 s_m	
R(F/S)=	F	-----			
	f_1	r_{11}	r_{12} r_{1m}	$R(f_1)$
	f_2	r_{21}	r_{22} r_{2m}	
	⋮			
	f_n	r_{n1}	r_{n2} r_{nm}	$R(f_n)$
		$R(s_1)$	$R(s_2)$ $R(s_m)$	

Cette matrice $R(F/S)$ nous donne la répartition des ressources utilisées par chaque fonction f_i : $R(f_i)$ et par chaque division : $R(s_j)$
 D'autre part si nous mesurons les flux d'information d'entrée I_e et de sortie I_s et les volumes de travaux V_{ij} à traiter pour chaque division et pour chaque fonction on peut déterminer encore les matrices $I(F/S)$ et $V(F/S)$

		S			
		s_1	s_2 s_m	
F		-----			
f_1		I_{e11} / I_{s11}	I_{e1m} / I_{s1m}	I_{e1} / I_{s1}
f_2				
⋮				
f_n		I_{en1} / I_{sn1}	I_{enm} / I_{snm}	I_{en} / I_{sn}
		I_e^1 / I_s^1	I_e^m / I_s^m	

S				
F	s ₁	s ₂s _m	
f ₁	V ₁₁	V ₁₂V _{1m}	V(f ₁)
f ₂	V ₁₂	V ₂₂V _{2m}	⋮
⋮				⋮
⋮				⋮
⋮				⋮
f _n	V _{n1}	V _{n2}V _{nm}	V(f _n)
	V(S ₁).....V(S _m)			

On peut comparer à l'aide de ces matrices les quantités d'informations traitées avec les ressources utilisées et de là déterminer ce qui ne fonctionne pas bien dans le système de gestion .

1-2 - Analyse Syntaxique :

Chaque système de gestion travaille en utilisant son fonctionnement un langage . Nous appellerons ^{langage} du système l'ensemble des chaînes de symbole qui décrit les états d'entrée et de sortie du système et de ses transformations. Les chaînes de ce langage sont la base de prise de décisions. L'analyse syntaxique consiste en l'analyse de ce langage de gestion, à la recherche d'une structure rationnelle et une morphologie des termes et des expressions de ce langage, le plus intéressant serait de construire des phrases décrivant les états de l'objet de gestion et leur réflexion en décisions à prendre :

$$L_x \times L_q \longrightarrow R$$

- L_x : l'ensemble des phrases caractérisant les entrées du système

- L_q : l'ensemble des phrases décrivant les états du système

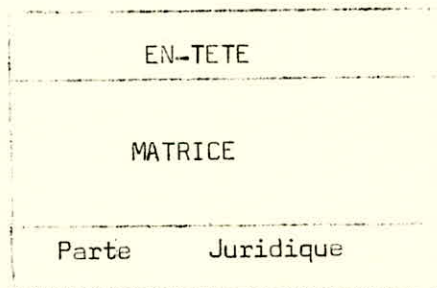
-R : Décisions

Les résultats de l'analyse du langage sont importants pour la réorganisation du traitement de l'information, la construction des fichiers, la distinction des données variables et des données constantes et pour l'intégration de tous les sous-systèmes déterminés pendant l'analyse logique de l'entreprise.

Comme nous l'avons précisé dans l'introduction le support de l'information qui circule dans un système de gestion sont les documents. On distingue dans les documents utilisés dans un système de gestion trois parties :

- Partie en-tête qui donne le nom du document
- Partie matricielle qui donne les caractéristiques principales de l'information véhiculé par ce document .
- Partie juridique: où il y a la signature et l'origine de leur provenance

On analyse essentiellement la partie matricielle qui contient l'information principale du document :



Ces documents qui contiennent une base informationnelle utilise un sous-ensemble du langage naturel : langage de l'information économique ou LIE. Ce langage a des caractéristiques remarquables telles que :

- Un lexique réduit
- Une syntaxe rigide mais simple et qui définit les liaisons entre les indicateurs du document.
- Le lexique du LIE a une redondance ou synonymie
- Une forme d'écriture en tableaux qui n'est pas commode pour le traitement automatique.

D'autre part il y a beaucoup de formes de documents qu'il faut unifier pour que l'analyse de ces documents puisse se faire facilement. Pour ce faire, on a distingué les informations qualitatives et des informations quantitatives dans la partie textuelle des documents qu'on a classé en différentes catégories: objets et prédicats, par la suite tirer les listes des expressions, des objets et des prédicats qu'on a classé suivant les relations qu'on établit dans le chapitre III.

1-3- ANALYSE SEMANTIQUE :

La sémantique était primitivement l'étude du sens des mots. Ce terme a pris, ces dernières années une telle extension que l'on se trouve parfois dérouté devant les emplois qu'on en fait. Ce terme vient du mot grec " Sēmainō " qui veut dire signifier, lui même dérivé de " Sēma ", c'est à dire signe . Or le dictionnaire donne pour " signifier " l'explication suivante :

avoir le sens de , être le signe de.

Donc la sémantique étudie le sens et le signe de l'information .

L'application la plus connue qui a été faite de la sémantique concerne la documentation scientifique. A ce propos on peut citer l'ouvrage par R.C GROSS, J.C GARDIN, F. LEVY et qui s'intitule (L'automatisation des recherches documentaires - un modèle general : le SYNTOL).

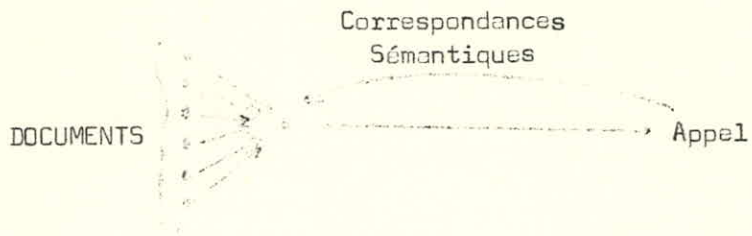
Pour ce qui nous concerne, on entend par analyse sémantique, l'analyse qui tient compte du volume, du sens et de l'utilité de l'information. On analyse le contenu et la structure de l'information contenue dans le document, on en tire deux listes : Une liste de noms ou objets et l'autres de prédicats. On peut construire à partir de ces listes des propositions atomaires qui sont des compositions de noms **et** de prédicats et la conjonction de ces propositions donne une description de l'état du système.

La valeur du contenu des phrases du langage sera donc le nombre de propositions atomaires, qu'il contient.

L'ensemble des termes forme un lexique ou dictionnaire, où la place de chaque descripteur d'entrée indique certains rapports sémantiques qu'il entretient avec les descripteurs de sortie.

Contrairement au langage documentaire où les données qui sont les indications portées sur les documents ne constituent pas toute l'information qu'il est utile d'enregistrer, car il existe des relations implicites qui doivent elles-même être enregistrées. Donc au lieu de simples listes de termes, on doit avoir une organisation sémantique du vocabulaire destinée à fixer le champ de regroupement des termes qu'il faut prévoir . Le langage de l'information économique contient toute l'information qu'il faut enregistrer, car les relations sémantiques apparaissent dans les documents entre les descripteurs .

Cette composition d'informations doit se trouver dans un ouvrage de référence ou " THESAURUS " . Un Thésaurus est un système informationnel à partir duquel il est possible de construire un traitement rationnel de l'information car les informations ont une organisation sémantique. Cette organisation sémantique nous facilite la recherche des informations nécessaires à un appel au système et nous permet d'établir la correspondance sémantique entre la formulation de la question (Appel) et le contenu des documents.



--- CHAPITRE II ---

II-1- MORPHOLOGIE DE L'INFORMATION ECONOMIQUE

Le support principal de l'information qui circule dans l'entreprise est le document . Un document se caractérise par :

- Sa partie ~~textuelle~~
- Sa structure

La partie textuelle se compare des descripteurs qu'on peut tirer et en faire une liste pour les analyser. Ces descripteurs véhiculent deux types d'informations élémentaires

- Les informations qualitatives qu'on représente par des t_i : ou expression

e_j

Exemple :

«terme» ::= Nom/Prénom/Ville/Rue/Mois/An/.....

«expression» ::= " terme"/"terme""expression"

expression: commencement du travail/Chiffre d'Affaires/Prix d'achat...

- Les informations quantitatives qu'on désigne par q_k : ce sont les significations des termes ou des expressions, elle peuvent être aussi des chiffres. Donc pour chaque terme ou expression, il existe une correspondance : q_k

Exemple :

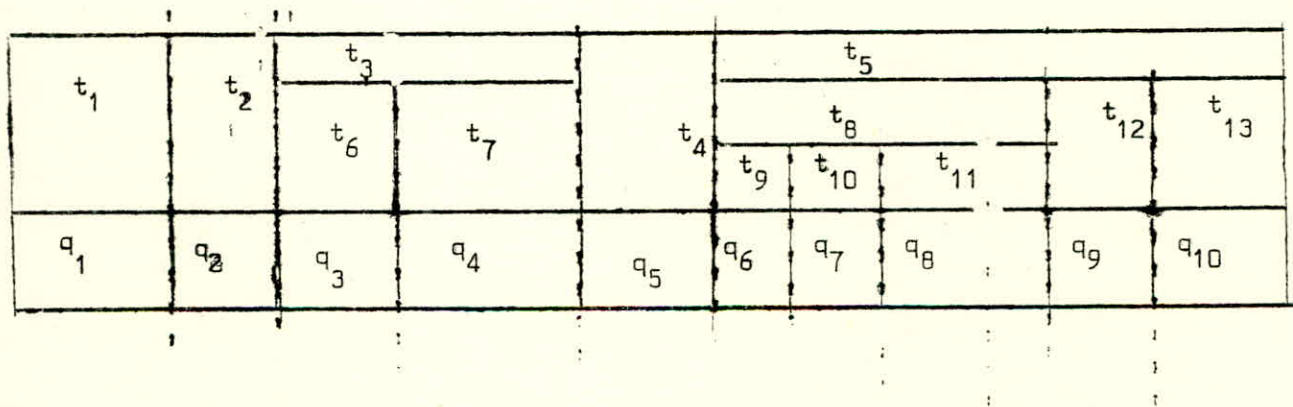
Nom - ALI , Somme -100DA, Commencement du travail -8h30,.....

De cette façon, on peut associer à chaque terme un ensemble de significations.

$$t_i \longrightarrow (q_{i1}, q_{i2}, \dots, q_{in})$$

La structure d'un document peut-être présentée comme un graphe de relations entre les différents descripteurs qui le composent.

Par exemple, si on a un document dont la structure est de la forme suivante :

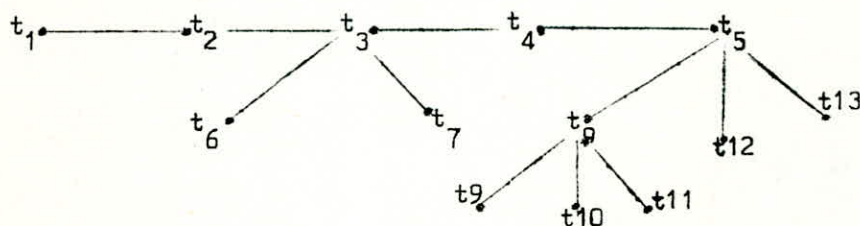


Dans cette structure, on a 10 descripteurs $d(q_i)$ auxquels on fait correspondre 10 significations $(q_1, q_2, \dots, q_{10})$.

La structure de ce document peut s'exprimer par les relations :

$$\begin{array}{ll} d(q_1) \longrightarrow t_1 & d(q_6) \longrightarrow t_5, t_8, t_9 \\ d(q_2) \longrightarrow t_2 & d(q_7) \longrightarrow t_5, t_8, t_{10} \\ d(q_3) \longrightarrow t_3, t_6 & d(q_8) \longrightarrow t_5, t_8, t_{11} \\ d(q_4) \longrightarrow t_3, t_7 & d(q_9) \longrightarrow t_5, t_{12} \\ d(q_5) \longrightarrow t_4 & d(q_{10}) \longrightarrow t_5, t_{13} \end{array}$$

Elle peut s'exprimer aussi par un graphe :



Elle peut s'exprimer aussi sous forme algébrique :

$$t_1 \cdot t_2 \cdot / t_3(t_6, t_7) / \cdot t_4 \cdot t_5 \cdot / t_8(t_9, t_{10}, t_{11}) \cdot t_{12} \cdot t_{13} /$$

D'autre part, les relations entre les termes dans cette structure obéissant à une grammaire du type :

$$\begin{array}{l} D \longrightarrow t_1 \cdot t_2 \cdot t_3 \cdot t_4 \cdot t_5 ; \\ t_3 \longrightarrow t_6 \cdot t_7 ; \\ t_5 \longrightarrow t_8 \cdot t_{12} \cdot t_{13} ; \\ t_8 \longrightarrow t_9 \cdot t_{10} \cdot t_{11} ; \\ t_1 \longrightarrow q_1 \\ t_2 \longrightarrow q_2 \\ \vdots \\ \vdots \\ t_{13} \longrightarrow q_{13} \end{array}$$

Dans ce cas, le vocabulaire terminal V_t représente l'ensemble des significations des termes ou des expressions : q_k

Les descripteurs jouent le rôle de vocabulaire non terminal : V_{nt} les règles de superposition $R : t_p \longrightarrow t_q, t_s, \dots, t_r$ sont définies par un graphe de relations.

Donc, le langage de l'information économique obéit à une grammaire qui est définie par:

- Un vocabulaire terminal : $V_t = (q_1, \dots, q_k)$
- Un vocabulaire non terminal : $V_{nt} = (t_1, t_2, \dots, t_n)$
- Un symbole initial D qui représente une phrase du langage
- Les règles de superposition R définissant les relations entre les termes

$$t_1, t_2, \dots, t_n.$$

II-2- Les ELEMENTS DE L'INFORMATION ECONOMIQUE :

Pour déterminer les éléments de l'information économique il faudrait analyser la partie textuelle et la structure du document .

La partie textuelle se compose des descripteurs : t_i, e_j dont on peut faire une liste. Cette liste sera une suite finie d'éléments e_j et t_i . Chaque expression de cette liste peut-être décomposée en deux parties principales :

- Une partie objet θ_i

un objet est un élément de l'information qui appelle à une description ou à une signification :

- Une partie prédicat qui explique l'objet c'est donc une partie explicative de la partie objet.

Exemple:

Soit l'expression :

' Nombre total de passagers par relation par vol par jour '

La partie objet est : nombre total de passagers

La partie prédicat est : relation, vol, jour

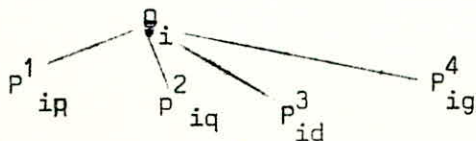
On peut distinguer dans une expression un certain nombre de prédicat dont les plus importants sont :

- Prédicat de temps : P_{ip}^1
- Prédicat de lieu : P_{iq}^2
- Prédicat de destination : P_{id}^3
- Prédicat divers : P_{ig}^4

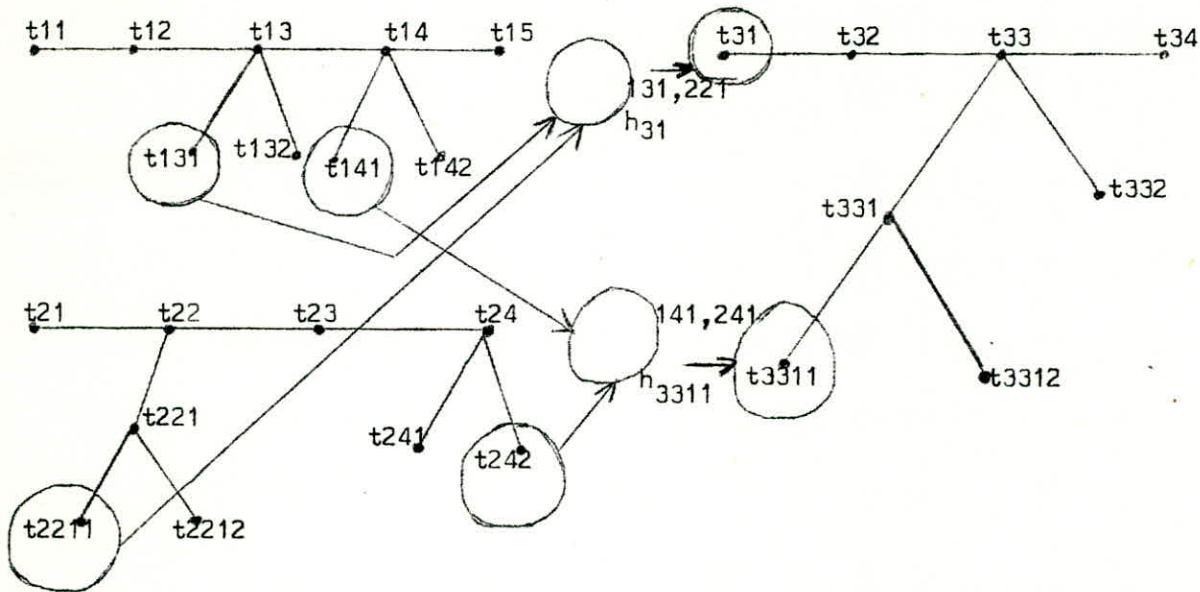
La forme générale d'une expression peut être représentée par :

$$e_j = \theta_i \cdot P_{ip}^1 \cdot P_{iq}^2 \cdot P_{id}^3 \cdot P_{ig}^4$$

On peut la représenter aussi sous la forme arborescente :



L'analyse de la structure des documents nous permet de distinguer les documents d'entrée, de sortie et les liaisons sémantiques qui les lient.



D'après les structures de ces documents on peut écrire :

$$d(t_{31}) = h_{31}^{131,221} / d(t_{131}), d(t_{221}) /$$

$$d(t_{3311}) = h_{3311}^{141,241} / d(t_{141}), d(t_{241}) /$$

On peut dire que t_{31} et (t_{131}, t_{221}) d'une part, et t_{3311} et (t_{141}, t_{241}) d'autre part ont des liaisons sémantiques.

On peut donc organiser un système de traitement de ces documents en appliquant aux descripteurs le principe des associations sémantiques.

- II-3 CLASSIFICATION - RELATIONS LOGIQUES :

On définit un dictionnaire pour les objets qu'on peut classer par ordre alphabétique et un autre pour les prédicats en définissant les relations hiérarchiques entre eux. Par exemple les prédicats du temps ont une structure hiérarchique car :

Année > trimestre > mois > semaine > jour

Dans un dictionnaire, chaque type de prédicat doit avoir une structure arborescente. En haut de l'arbre se trouve un prédicat général (année, Université...) puis les prédicats qui le détaillent (mois, Faculté...)

Prenons par exemple le descripteur : " Charge d'études des professeurs du département pour le mois "

- Ce descripteur Groupe

- un objet : charges des professeurs
- un prédicat de temps : mois
- un prédicat de niveau : département
- un prédicat divers : études

Les relations entre les prédicats influent sur les relations entre les descripteurs et sur les procédures de leur traitement car il y a des relations logiques entre les prédicats et les descripteurs :

- relation d'inclusion (entier - partie) R_{ep} :

Si nous avons un objet Θ tel que $P_1 \subset P$ cela signifie qu'un prédicat P_1 de l'objet Θ appartient au prédicat P de Θ . La classe de tous les prédicats inclus est un ensemble fini des significations du prédicat donc ils sont une partie :

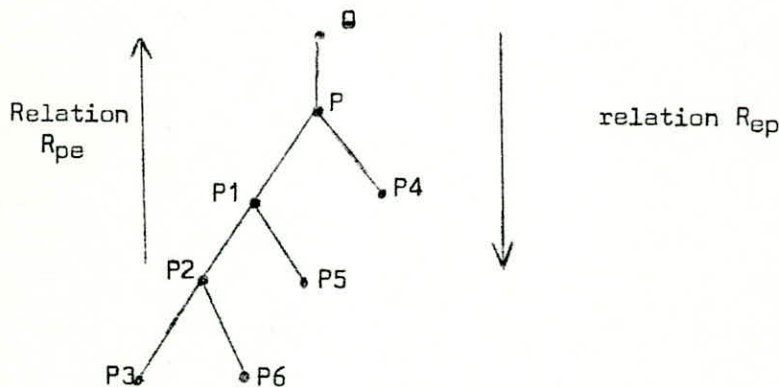
Si $P \supset P_1$; $P \supset P_2, \dots, P \supset P_L$ $P = \cup_{i=1}^L P_i$

La valeur d'un prédicat peut être trouvée dans l'arbre d'inclusions car :

$$\forall i, j, P_i \cup P_j = P_{ij}$$

- relation d'entrée (partie-entier) R_{pe}

C'est une relation opposée à la précédente. Elle permet de trouver un prédicat dans l'arbre des prédicats mais dans une direction inverse à la précédente c'est à dire en remontant dans l'arbre des prédicats.



- relation d'équivalence (synonymie) R_S

Si nous avons deux termes ou descripteurs T_1 et T_2 qui ont une relation sémantique R_S avec T_3 et T_4 tel que :

$$T_4 = H_1(T_1, T_3) = H_1(T_2, T_3) \implies T_1 \equiv T_2$$

Dans le dictionnaire des synonymes, on utilisera la relation R_S

- relation sémantique (correspondance sémantique) R_{Sm}

C'est une relation définie de la façon suivante:

$$\forall t_j \exists (t_1, t_2, \dots, t_k) \text{ tel que } t_j = h_i^{1, \dots, k}(t_1, t_2, \dots, t_k)$$

Avec t_j = description de sortie (nécessaire pour la prise de décisions, et
et pour l'analyse de la gestion)

t_1, t_2, \dots, t_k : descripteurs d'entrée.

2-4 Modèle logique de l'information économique

Le langage de gestion est destiné à exécuter deux fonctions =

- recherche de l'information
- traitement de l'information

Pour cela, le langage de l'information économique doit être organisé
de telle façon qu'il ait :

- un dictionnaire des termes (objets et prédicats)
- un dictionnaire spécial qui contient des liaisons sémantiques entre les
descripteurs.
- une composition de procédures standards de traitement
- un dictionnaire des synonymes.

Le dictionnaire des termes peut être construit automatiquement. Pour cela
on introduit la liste des expressions qu'on aura auparavant, tirée des documents
utilisés et avec indication préalable des objets Q_j . Un programme pourra alors tirer
les termes et les classer par ordre alphabétique et éventuellement les codifier.

Le dictionnaire des prédicats est aussi tiré de la liste des expressions
avec dans ce cas une classification par type de prédicat et avec l'indication des
relations entre les prédicats de chaque type.

Dans une expression on trouve généralement tous les descripteurs et la
forme générale d'une expression est :

objet Q P_{i1}^1 P_{i2}^2 P_{id}^3 P_{ig}^4

Mais pour une recherche rapide des termes et des expressions dans un
dictionnaire, il conviendrait de codifier auparavant tous les mots : comme par exemple :

A	B	C	D
---	---	---	---

A = code de longueur

B = code de groupe

C = numéro d'ordre

D = code de l'objet ou prédicat

Une telle présentation est utile pour l'analyse de l'utilisation des descripteurs dans les documents. Car après avoir stocké les descripteurs des documents on fait un tri pour éliminer les répétitions des descripteurs avec le calcul des fréquences de leur utilisation on aura alors en sortie :

code terme fréquence d'utilisation

d'autre part, avec les dictionnaires des termes et des synonymes. Nous pouvons alors organiser la recherche puisqu'on a un code par mot,

Les descripteurs des documents ont une forme différente, une séquence différente des termes etc.... Pour un traitement rationnel, il faut standardiser la forme des descripteurs. On peut les présenter sous forme canonique de la façon suivante :

code de l'objet code de P_{i1}^1 code de P_{i2}^2 code de P_{id}^3 code de P_{ig}^4

Cette forme canonique nous permet de présenter tous les descripteurs sous une forme unique.

Quand on a une phrase en entrée d'un système nous pouvons l'analyser, trouver les codes de tous les termes et avoir alors sa description complète.

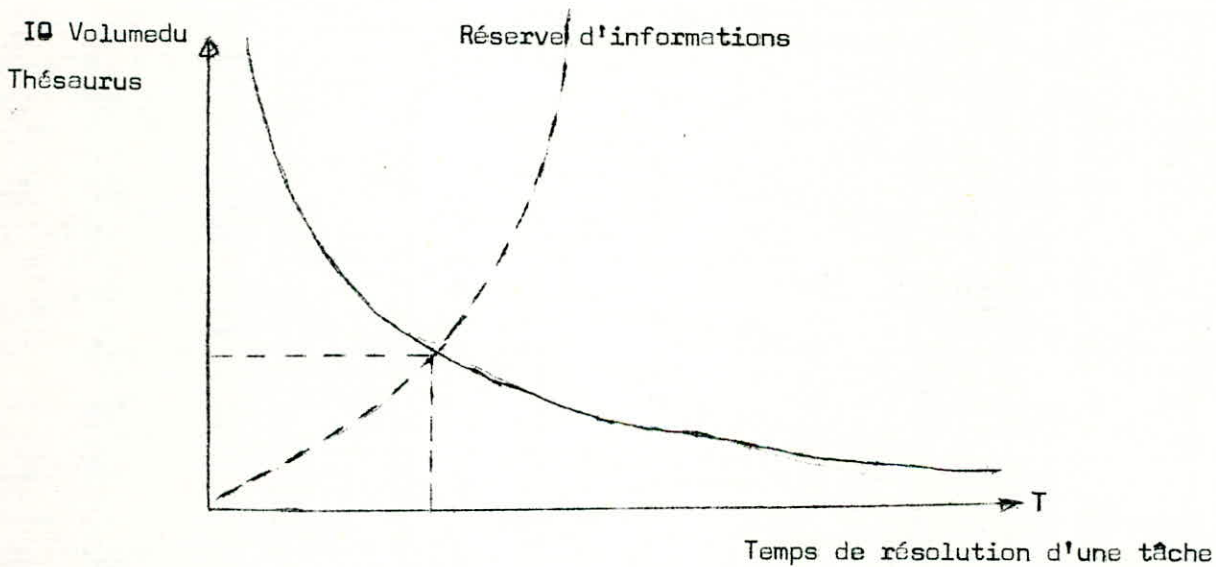
Après avoir construit les dictionnaires des objets prédicats on peut construire un tableau des liaisons paradigmatiques entre les objets et les prédicats comme il sera explicité plus tard, ce tableau nous indique l'organisation des données dans le système de traitement sémantique :

	P	P_1^1	P_2^1 P_p^1	P_f^2 P_q^2	P_1^3 P_d^3	P_1^4 P_g^4
θ_1	1	0	1.....0	1....1	1....0	0.....0
θ_2	1	1	0.....1	0....0	0...0...0.....1	
\vdots	\vdots					
θ_n	1	0	0.....0	0....1	1... 0..1.....1	

L'indication paradigmatique traduit certains rapports essentiels, généralement stables et implicites qui existent entre les termes d'un document.

2-5- Notions De Thésaurus :

Chaque receveur d'information a toujours un ensemble d'informations préalables qui constituent sa réserve d'information et la réception d'une information nouvelle modifie le contenu de sa réserve d'informations. Si cette composition d'informations est consignée dans un ouvrage avec en plus l'indication des relations sémantiques de ces informations on aura alors constitué un thésaurus du receveur. La quantité et la qualité des informations du thésaurus définissent la structure et le volume de celui-ci. Aussi un système avec un thésaurus développé peut recevoir plus d'information et résoudre une tâche donnée plus vite qu'un système avec un thésaurus moins développé. Aussi le volume du thésaurus détermine la capacité du receveur.



La construction d'un thésaurus doit faciliter la recherche des descripteurs nécessaires pour répondre à un appel. Aussi par rapport à une banque de données la recherche dans un thésaurus est plus rapide car chaque élément du thésaurus est défini par son code et ses liaisons sémantiques avec les autres éléments. Ces liaisons sémantiques se reflètent dans un tableau de liaisons. Ce n'est pas le seul avantage qu'a le thésaurus sur la banque de données. En effet :

- il occupe moins de place dans la machine qu'un fichier
- il est plus simple et reflète les relations existants entre les descripteurs
- il est flexible et permet donc l'étude de systèmes de différentes catégories
- il est ouvert et permet donc le développement d'un système
- il permet de typiser les descriptions des documents.

On peut citer trois méthodes principales pour construire un Thésaurus:

1) Méthode directe:

Pour chacun des m documents, on définit les descripteurs caractéristiques de chaque document et on construit une matrice $(m \times n)$ pour définir leurs relations.

descripteurs	d1	d2	d3.....	dn
documents				
D1	0	1	1.....	0
D2	1	0	1.....	0
D3
.
.
.
Dn	0	0	1.....	1

$$a_{ij} = (0, 1)$$

Ainsi si dans l'appel figure les descripteurs d_2 et d_3 on aura le document D_1 : $D_1 = d_2 \cup d_3$

2) Méthode inverse:

Pour chacun des n descripteurs on définit la liste des documents associés ou des autres descripteurs associés. Ainsi la méthode inverse contrairement à la méthode directe l'image informationnelle sont les documents associés aux descripteurs.

documents descripteurs	D1	D2	D3.....	Dn
d1	1	0	1	1
d2	1	1	1	0
.
.
.
dn	1	1	0	1

Ainsi les documents D1, D2 et D3 ont en commun les descripteurs d1, d2 et dn :

$$d1 = D1 \cap D3 \cap Dn$$

$$d2 = D1 \cap D2 \cap D3$$

$$dn = D1 \cap D2 \cap Dn$$

Aussi si dans l'appel figure le descripteur : d2 on aura les documents : D1, D2, D3.

3) Méthode duale:

Cette troisième méthode utilise les deux précédentes. On construit un tableau de descripteurs ordonnés et pour chaque descripteur on indique tous les documents qui le contiennent.

	documents			descripteurs			
	D1	D2	Dn	d1	d2	d3.....	dn
d1	1	0	1	1	0	1.....	1
d2	0		1	1	0	1.....	0
.
.
.
dn	0	0	1	0	1	1	0

Pour choisir un moyen de construction de Thésaurus il faudrait analyser le système existant et ce, sous formes :

- 1) Analyse systématique pour définir le but du système existant et pour pouvoir décomposer le système en sous-système avec leurs fonctions.
- 2) Analyse fonctionnelle pour déterminer les tâches correspondants et définir pour chacune d'entre elles les informations d'entrée et de sortie.

3) Etape de construction des algorithmes de réception de chaque indicateur. Et après suivant: le volume des documents, le volume moyen des images informationnelles, la **fréquence** de changement des informations du Thésaurus, la composition des appels. On choisit l'une des méthodes de construction du Thésaurus.

Dans notre cas, il est possible de construire le Thésaurus pour la Compagnie Air-Algérie en appliquant la méthode inverse. En effet, après avoir analysé les documents et déterminé tous les descripteurs on constate qu'il est plus facile de construire le Thésaurus à l'aide de cette méthode car le nombre de documents est faible par rapport au nombre de descripteurs.

Le cas concret qu'il nous a été possible de considérer est celui de la compagnie nationale AIR-ALGERIE. Vu le développement que connaît cette compagnie par l'extension de son réseau, le nombre de documents qu'elle a à traiter a augmenté très vite. Or, le traitement de ces documents est essentiellement manuel, aussi cette compagnie au niveau de la direction des recettes commerciales accuse généralement un retard de six mois dans le traitement des documents. Ce retard se répercute au niveau de l'étude de l'évolution de l'entreprise, ainsi qu'au niveau des prises de décisions.

Pour améliorer ce traitement on va essayer de construire un système sémantique, aussi comme nous l'avons souligné dans les précédents chapitres il est nécessaire de faire l'analyse des documents pour tirer la liste des expressions et celle des objets et des prédicats.

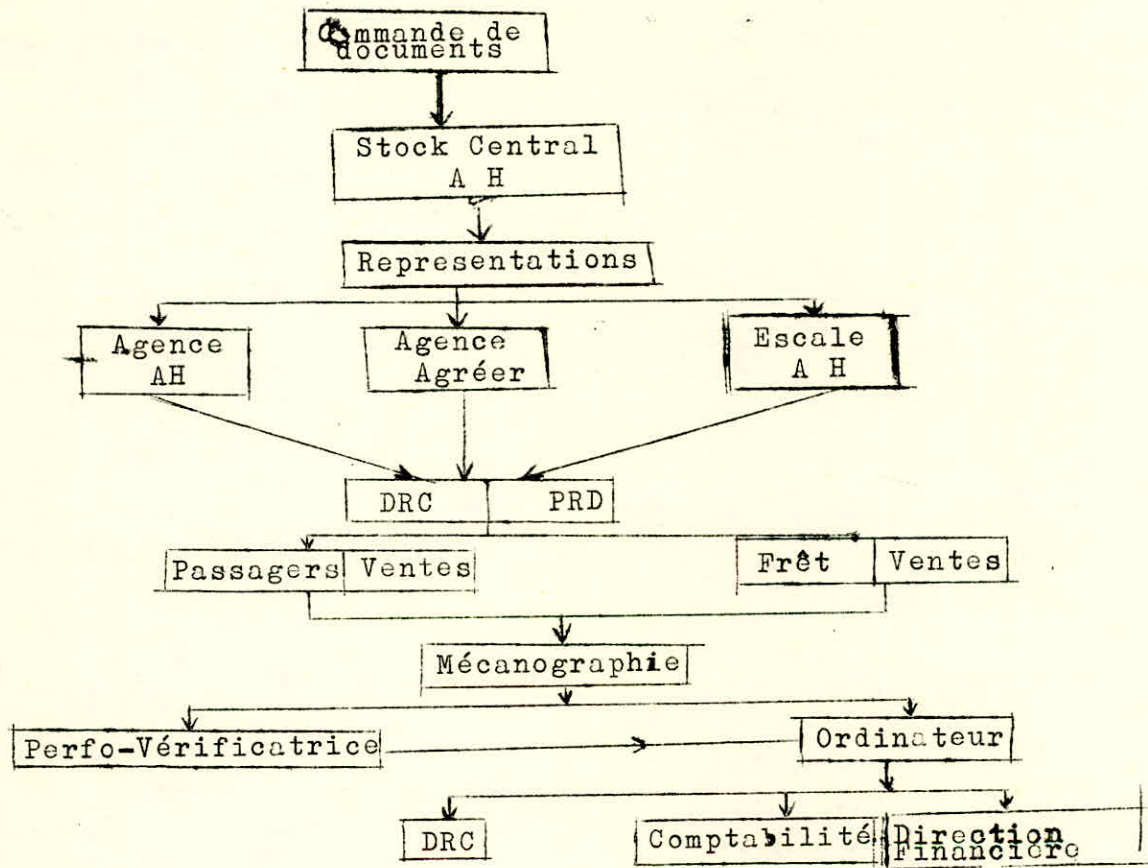
3-1 : Chaines de traitement AIR-ALGERIE

La direction des recettes commerciales reçoit trois enveloppes des différentes agences du pays, de ses représentations à l'extérieur et des escales

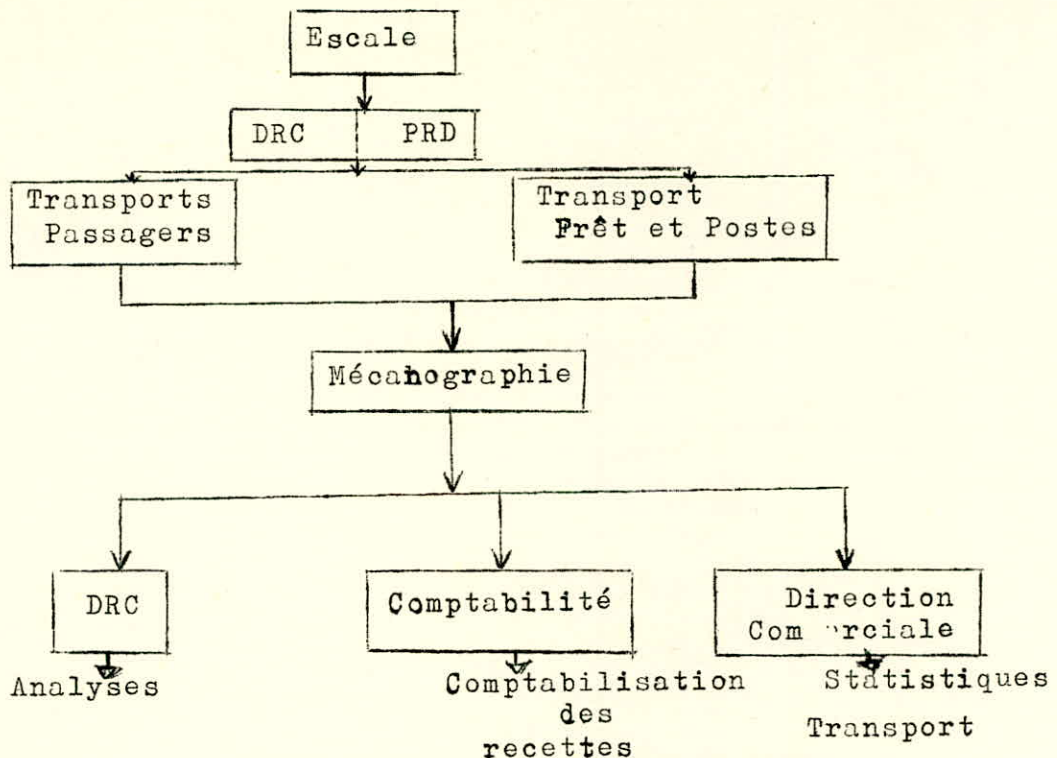
- une enveloppe "transport" (coupons de vol, BCB, EBV)
- Une enveloppe " Frêt " (LTA)
- Une enveloppe " Poste " (AV7)

Le traitement de ces enveloppes consiste en une vérification des différents documents, vérification qui consiste en l'exactitude des sommes portés et dans le cas d'erreurs à émettre des réclamations au point d'émission des documents. L'autre traitement consiste à calculer les différentes valeurs nécessaires pour suivre l'évolution de la compagnie et à prendre les décisions qui s'imposent.

Nous n'avons pu traiter que le cas du transport, vu la difficulté qu'on a eu pour recueillir les documents sur les deux autres cas.



Organigramme : Circuits des documents



CIRCUIT DES DOCUMENTS TRANSPORT :

La première phase de notre travail a été la collecte des documents principaux relatifs à la section "transport", la seconde a consisté en l'analyse de ces documents.

La compagnie "AIR-ALGERIE" réalise la fonction globale F: Vente et transport. Cette fonction principale F se décompose en 3 sous-fonction f_i :

f_1 = Ventes

f_2 = Transport

f_3 = Frêt et poste

Concernant la sous-fonction f_2 (transport) on a dénombré six tâches Z_{mi}^1

Z_{21} = Contrôle fichier ligne-enveloppe

Z_{22} = Facturation reçue " Passagers "

Z_{23} = Facturation émise compagnies étrangères

Z_{24} = Transport " Passagers "

Z_{25} = Etat d'ajustement " Facturation reçue "

Z_{26} = Prix moyens " Passagers ".

Pour chacune de ces tâches on a trié les descripteurs de sortie et d'entrée de telle façon à pouvoir tirer la liste des objets et des prédicats pour enfin construire le Thésaurus de cette sous fonction " Transport ".

3-2- ANALYSE DES DOCUMENTS :

Il est à remarquer que nous n'avons pas tenu compte des documents intermédiaires qui sont nombreux mais qui n'entrent pas dans le traitement sémantique. Les documents dont nous avons tenu compte sont :

- Coupons de vol (C.V)
- Bulletin complémentaires de bagages (B.C.B)
- Etat des billets vendus (E.B.V)
- Mettre de transport aérien (L.T.A)
- AV7

Après cette collecte de documents on a établi une méthode d'enquête. Cette méthode consiste à déterminer les tâches de la fonction transport dont nous avons à établir le traitement sémantique. Ces tâches sont au nombre de 6 et pour chaque tâche nous avons à déterminer les descripteur d'entrée et de sortie ainsi que les relations sémantiques qui les lient.

Exemple :

Descripteurs de sortie	Descripteurs d'entrée associés	Realations
T N B P	-NBAH	TNBP=NBAH + NBCE
T R E C	- REAH	TTREC=REAH+RECE
P M O Y	TREC	PMOY= TREC/ TNBR
	NBCE	
	RECE	
	TNBP	

TACHE I

CONTROLE FICHER -LIGNES ET ENVELOPPES

DOCUMENTS D'ENTREE : CV, BCB, AH ET CE

DESCRIPTEURS D'ENTREE		DESCRIPTEURS DE SORTIE	RELATIONS
CODE ENREGISTREMENT	: CEN	CODE ENREGISTREMENT	: CEN
SECTEUR TRANSPORT	: STR	SECTEUR TRANSPORT	: STR : STR=STR
DATE CREATION	: DCR	DATE CREATION	: DCR : DAT=DAT
N° LIGNE	: NLIG	N° LIGNE	: NLIG : DIST=DIST
NATURE VOL	: NAV	NATURE VOL	: NAV : COF=CHARG
TYPE APPAREIL	: TYAP	TYPE APPAREIL	: TYAP : FR=FR
CODE POOL	: POOL	DISTANCE	: DIST : PMPAP1=
RELATION NUMERIQUE	: RELN		: PMPAP1
DUSTANCE	: DIST		: PMPAP2=
CHARGE OFFERTE	: COF	CHARGE	: PMPAP2
FREQUENCE	: FR	FREQUENCE	: CHARG : PMFRET=
PRIX MOYEN PAPI	: PMPAP1		: FR : PMFRET
PRIX MOYEN PAP2	: PMPAP2	PRIX MOYEN PAPI	: FR : PMPAP1=
PRIX MOYEN FRET	: PMFRET		: PMPAP2 : PMPOST
PRIX MOYEN POSTES	: PMPOST	PRIX MOYEN PAP2	: PMPAP1 : SOFPAP1
SIEGES OFFERTS PAPI	: SOFPAP1		: PMPAP2 : SOFPAP1
SIEGES OFFERTS PAP2	: SOFPAP2	PRIX MOYEN FRET	: PMPAP2 : SOFPAP2
DATE	: DAT	PRIX MOYEN POSTES	: PMPOST
REALATION ALPHA	: RELA	SIEGES OFFERTS PAPI	: SOFPAP1
N° ENVELOPPE	: NENV	SIEGES OFFERTS PAP2	: SOFPAP2
NATURE DOCUMENT	: NAD	DATE	: DAT
		RELATION	: REL

Nous faisons la correspondance entre les lignes et les enveloppes pour s'assurer que les enveloppes recues proviennent bien d'un vol programmé

PREDICATS : N° LIGNE, NATURE VOL, RELATION NUMERIQUE, TYPE APPAREIL, RELATION ALPHA

TACHE III :

STATISTIQUES FACTURATION 'EMISE' (C.E)

DOCUMENTS D'ENTREE

FACTURATION EMISE PAR AH

DESCRIPTEURS D'ENTREE	DESCRIPTEURS DE SORTIE	RELATIONS
RELATION	REL : N° AGENCE	NAG : NBCE = $\sum C.VPCE$
N° LIGNE	NLIG : RELATION	REL :
NATURE L VOL	NAV : NOMBRE PASSAGERS PAYANTS	NBCE : RECE = $\sum TOT C$
TYPE APPAREIL	TYAP : RECETTES PASSAGERS PAYANTS	RECE : $PMOY = \frac{RECE}{NBCE}$
CODE BILLET	CB : PAX GRATUIT	PGR :
BASE TARIF	BATA : SERVICE	SER :
TRANSPORTEUR	TRANS : BEBES	BEBE : TOTAC = $\sum TOTAC$
COMPAGNIE EMETTRICE	COMEM : PRIX MOYEN	PMOY : TOTGR = PGR + SER + BEBE
DATE VOL	DATV : TOTAL AGENCE F	TOTAGF : TOTAGF = $\sum C.V.F$
	TY : TOTAL AGENCE TY	TOTAGTY : TOTAGTY = $\sum C.V.TY$
	CLAF : TOTAL AGENCE TC	TOTAGTC : TOTAGTC = $\sum CVTC$
CLASSE TY	CLATY : TOTAL COMPAGNIE F	TOTAGF : TOTCF = $\sum TOTAGF$
CLASSE TC	CLATC : TOTAL COMPAGNIE TY	TOTGY
PRIX	PR : TOTAL COMPAGNIE TC	TOTCTS : TOTCTY = $\sum TOTAGTY$
TAXE	TAX : TOTAL GENERAL F	TOTGF : TOTCTC = $\sum TOTAGTC$
CONTRE VALEUR	COVA : TOTAL GENERAL TY	TOTGY : TOTGF = $\sum TOTCF$
TOTAL	TOT : TOTAL GENERAL TC	TOTGTC : TOTGTY = $\sum TOTCTY$
MODE DE PAYEMENT	MOPA : TOTAL GRATUITS	TOTGR : TOTGTC = $\sum TOTCTC$

PREDICATS : RELATION, N° LIGNE, DATE, CLASSE

TACHE IV

TRANSPORT PASSAGERS

DOCUMENTS D'ENTREE : C-V , EBV

DESCRIPTEURS D'ENTREE	DESCRIPTEURS DE SORTIE	RELATIONS	
RELATION	'REL	'REL	'NBPAAH = $\sum UVAAH$
N° LIGNE	'NLIG	'NLIG	'REPAH = REAAH
NATURE VOL	'NAV	'NOMBRE PASSAGERS	'NBPCE = $\sum CVCE$
TYPE APPAREIL	'TYAP	'RECETTES PASSAGERS	'REPCE = REECE
DATE	'DAT	'ers	
SECTEUR	'SECT	'NOMBRE PASSAGERS	'NBT = NBPAAH + NBPCE
CODE POOL	'POOL	'RECETTES PASSAGERS	'RECT = REPAH + REPCE
NOMBRE PAYANTS AH	'NBAH	'NOMBRE TOTAL	'PMOY = $\frac{RECT}{NBT}$
RECETTES PAYANTS AH	'REAAH	'RECETTES TOTALES	
NOMBRE PAYANTS CE	'NBCE	'PRIX Moyen	'TOTMO = $\sum TOTJ$
RECETTES PAYANTS CE	'REECE	'TOTAL MOIS	
GRATUITS PO	'GRPO	'TOTAL RELATION	'TRE = $\sum RECT$
GRATUITS P1	'GRP1		
GRATUITS P2	'GRP2		
PGRATUITS PN	'GRPN		
GRATUITS FAVEUR	'GRFA		
GRATUITS BEBES	'GRBE		
GRATUITS SERVICE	'GRSE		

PREDICATS : RELATION - N° LIGNE - NATURE VOL - TYPE APPAREIL

TACHE V :

ETATS D'AJUSTEMENT FACTURATION RECUE

DOCUMENTS D'ENTREE

FACTURATION RECUE

- C V -

DESCRIPTEURS D'ENTREE	DESCRIPTEURS DE SORTIE	RELATIONS
N° ENVELOPPE	NENV : N° COMPAGNIE	NCOM :
N° DOCUMENT	NDOC : N° FACTURE	NFACT :
DATE	DAT : CODE MONNAIE	C M :
N° FACTURE	NFACT : N° ENVELOPPE	NENV : MMF=MMEV=CP
CODE MONNAIE	C M :	
RELATION	REL : N° DOCUMENT	NDOC :
NATURE UTILISATION	NUTL : MONTANT MONNAIE	MMF : COM=PR.PCO
CLASSE F	CLAF : FACTURATION	VADOC : VADOC= PR
CLASSE TY	CLATY : VALEUR DOCUMENT	COM :
CLASSE TC	CLATC : COMMISSIONS	
N° C.V	NCV :	
N° COMPAGNIE	NCOM :	
N° AGENCE	NAG :	
MONTANT MONNAIE	MMEV :	
EVALUATION		
TAXES	TAX :	
BASE TARIF	BATA :	
% COMMISSIONS	PCO :	
CODE POOL	CP :	
PRIX	PR :	

PREDICATS : N° COMMANDE - N° FACTURE = N° ENVELOPPE + RELATION

TACHE VI

PRIX MOYEN PASSAGERS RELATION / LIGNE

DOCUMENTS D'ENTREE : TITRE DE TRANSPORT (C.V) AH et CE

DESCRIPTEURS D'ENTREE		DESCRIPTEURS DE SORTIE	RELATIONS
RELATION	' REL	RELATION	' REL ' NBAH = Σ CVP AH
N° LIGNE	' NLIG	N° LIGNE	' NLIG ' REAH = Σ TOTAH
NATURE VOL	' NVOL	NOMBRE PAYANTS	' SNBAH ' NBCE = Σ CVPCE
TYPE APPAREIL	' TYAP	RECETTES A H PAYANTS A H	' REAH ' RECE = Σ TOTCE
CODE BILLET	' CB	NOMBRE PAYANTS	' NBCE ' TNBP = NBAH + NBCE
BASE TARIF	' BATA	RECETTES PAYANTS	' RECE ' TREP = REAH + RECE
TRANSPORTEUR	' TRANS	CE NOMBRE TOTAL PAYANTS	' TNBP ' NBGR = Σ CVGR
COMPAGNIE EMETTRICE	' COMEM	RECEPTE TOTAL PAYANTS	' TREP ' PMOY = $\frac{TREP}{TNBP}$
DATE VOL	' DATV	PASSAGERS GRAB tuits	' PAGR
CLASSE F	' CLAF	PASSAGERS PO	' PAPO
CLASSE TY	' CLATY	PASSAGERS P1	' PAP1
CLASSE TC	' CLACT	PASSAGERS P2	' PAP2
PRIX	' PR	PASSAGERS PN	' PAPN
TAXES	' TAX	PASSAGERS FAVEUR	' PAPA
CONTRE VALEUR	' COVA	SERVICE S1	' SERS1
TOTAL	' TOT	SERVICE S2	' SERS2
MODE PAIEMENT	' MOPA	BEBES PRIX MOYEN	' BEBE ' PMOY
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'
	'	'	'

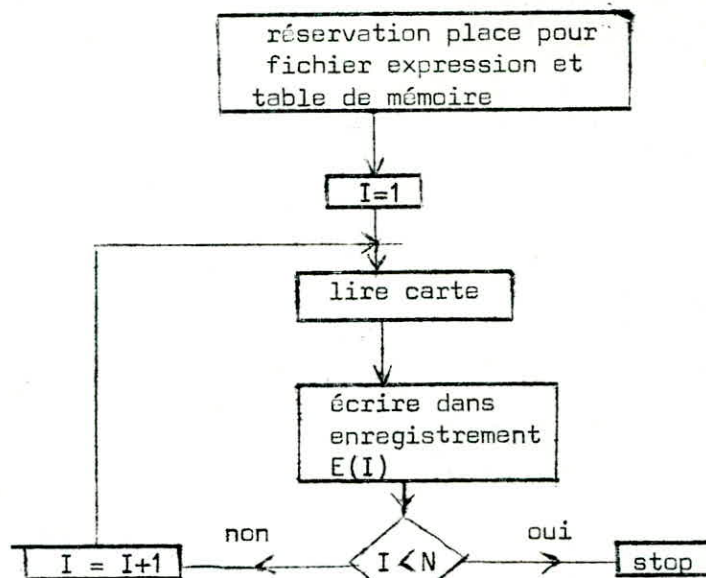
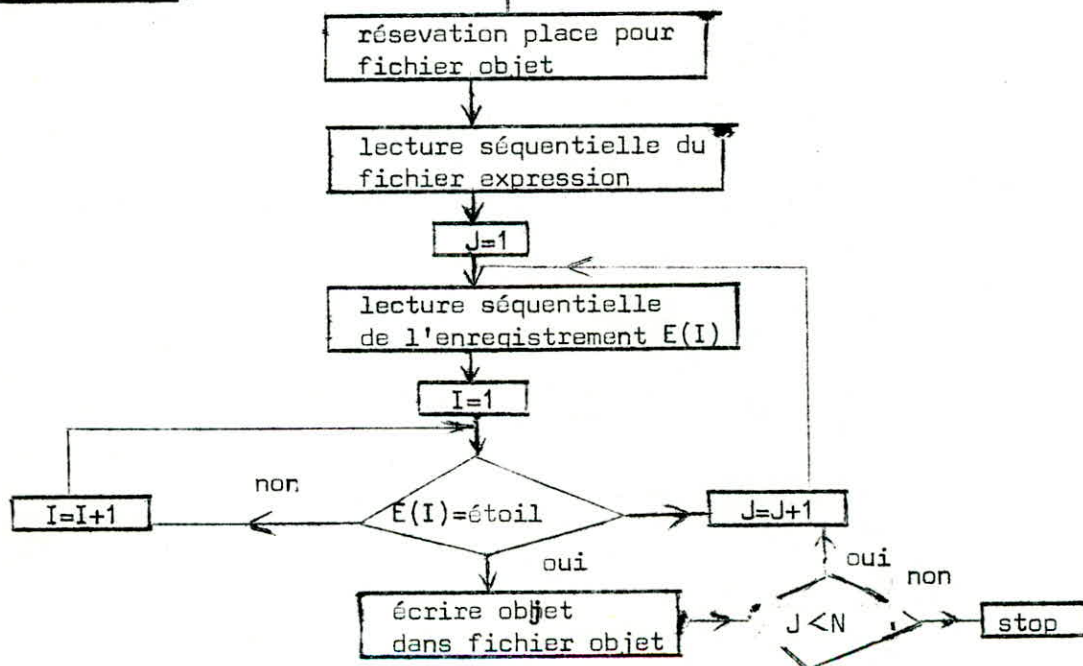
PREDICATS : RELATION - N° LIGNE - DATE VOL - CLASSE

- 1 _ Numero Secteur
- 2- Numéro ligne
- 3_ Numéro Transporteur
- 4- Numéro Compagnie emettrice
- 5- Numéro agence
- 6 - Relation numérique
- 7- Relation alphabétique
- 8- Nature Vol
- 9- Type appareil
- 10- Distance *
- 11- Fréquence *
- 12- Charge offerte *
- 13- Date Vol
- 14- Code Pool
- 15- numéro enveloppe
- 16- Numéro Document
- 17- Nature Document
- 18- Numéro billet
- 19- Base Tarif *
- 20- Prix *
- 21- Taxes *
- 22- Contre valeur *
- 23- Total à payer (billet) *
- 24- % Commissions *
- 25- Mode Paiement
- 26- Classe F
- 27- " " TY
- 28- " " TC
- 29- Gratuit Po
- 30- " " PI
- 31- " " P2
- 32- " " PN
- 33- " " Faveur
- 34- " " Bébé
- 35- " " Service SI
- 36- " " Service S2
- 37- Code monnaie

38- Numéro Facture
 39- Montant monnaie évaluation
 40- Montant monnaie * Facturation
 41- nombre*Pax payants
 42- Recettes * Pax Payants
 43- Nombre * Pax Payants
 44- Recettes * Pax Payants
 45- Total * Nombre * payants
 46- Total * Recettes * payants
 47- Prix moyen * PAP1
 48- " " * PAP2
 49- Prix moyen * frêt
 50- " " * Postes
 51- Sièges offerts * PAP1
 52- Sièges offerts * PAP2
 53- Nombre * Passagers Gratuits P0
 54- Nombre * " " " " P1
 55- " * " " " " P2
 56- " * " " " " PN
 57- " * " " " " Faveurs
 58- " * " " " " bébé
 59- " * " " " " service S1
 60- " * " " " " service S2
 61- Montant Commissions *
 62- Facturation globale *
 63- Total * Nature document
 64- Total * monnaie
 65- Total */Agence/classe F
 66- Total*/Agence/classe TY
 67- Total*/Agence/classe TC
 68- Total */Pays/classe F
 69- Total* /Pays/classe TY
 70- Total */Pays/classe TC
 71- Total*/Compagnie/Classe F
 72- Total* /Compagnie/Classe TY
 73- Total */Compagnie/Classe TC
 74- Total * Général
 75- Total * Gratuit

76- Total * / relation

77- Total * / mois

Fichier expressionsFichier objets

3-3 : Liste des objet et prédicats

de Ces tâches on tire la liste des expressions contenues dans toutes les tâches de cette liste des expressions tirer celle des objets et des prédicats. Les listes des objets et des prédicats sont établis sans répétitions des descripteurs.

A- Liste des objets :

- Distances
- Fréquence
- Charge offerte
- Base tarif
- Prix
- Taxe
- Contre-valeur
- Total à payer
- % Commissions
- Montant monnaie
- Nombre
- Recettes
- Total
- Prix moyen
- Siéges offerts
- Montant Commissions
- Facturations globale.

C- Liste des prédicats:

Cette liste de prédicats est à organiser en classes, nous avons distingué sept calsses de prédicats : Prédicat de place, prédicats de Compagnie, prédicat de classe, prédicats de paye, prédicat pool, prédicat de temps, prédicats de numéro.

1 prédicats place

AH
Secteur
Agence
Ligne
Relation

4 Prédicat paye

Mode de paiement
Pax payants
Cash
Crédit
Chéque
Code monnaie
Pax gratuits
Pax en

Pax P2
 Pax PN
 Pax bébés
 Pax ~~favor~~
 Pax service S1
 Pax service S2

2 Prédicat Compagnie

AH
 CE
 Transport
 émission
 numéro vol
 type appareil

3 Prédicat classe

classe
 frêt
 poste
 pax
 TY
 TC
 F
 PAP1
 PAP2

5 Prédicat pool

code pool
 évaluation
 facturation

6 Prédicat temps

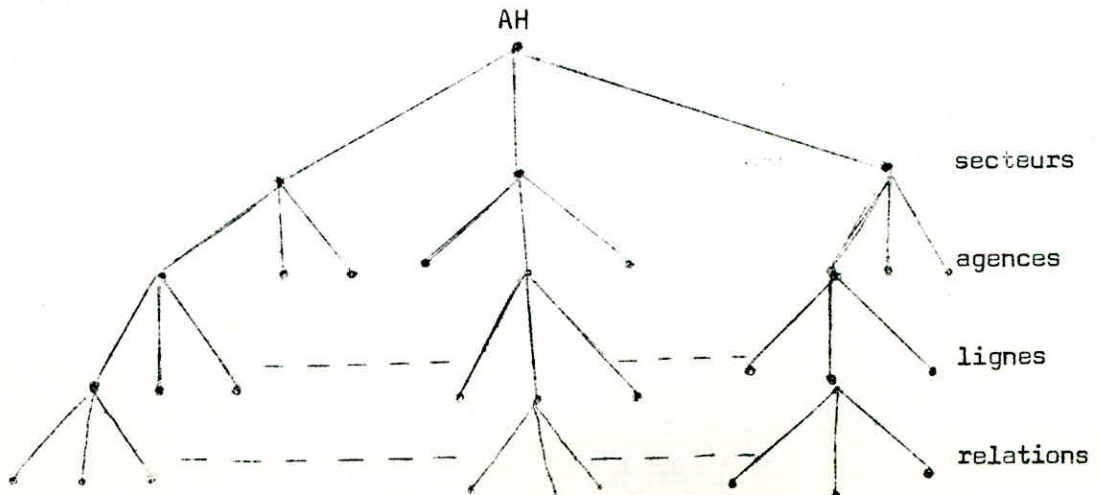
Date
 année
 mois
 semaine
 jours

7 Prédicat numéro

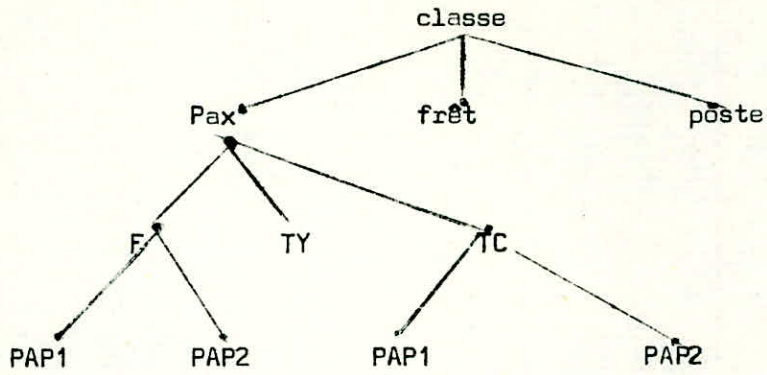
numéro enveloppe
 nature document
 numéro document
 numéro facture
 numéro billet

Comme nous l'avons indiqué précédemment, il nous faut établir maintenant les relations hiérarchiques des prédicats :

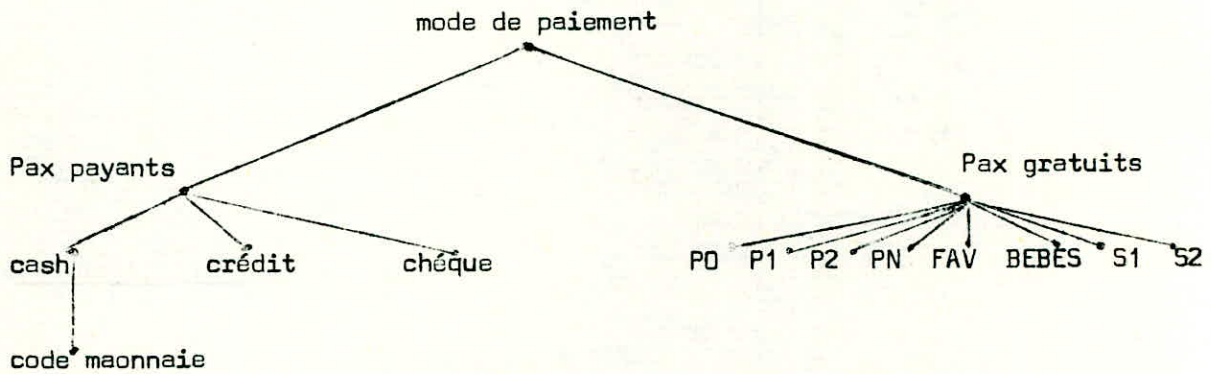
Prédicat Compagnie



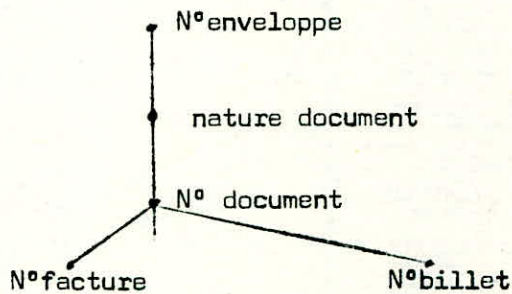
Prédicat classe



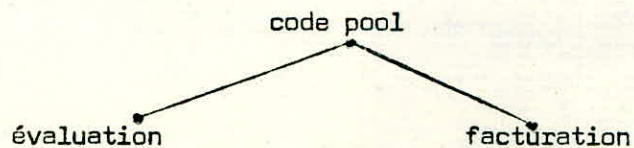
Prédicat paye

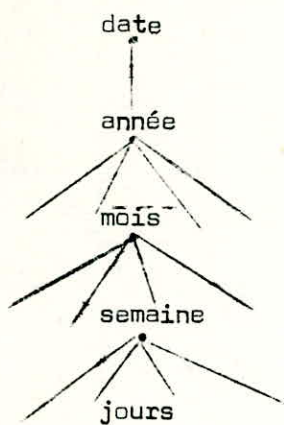
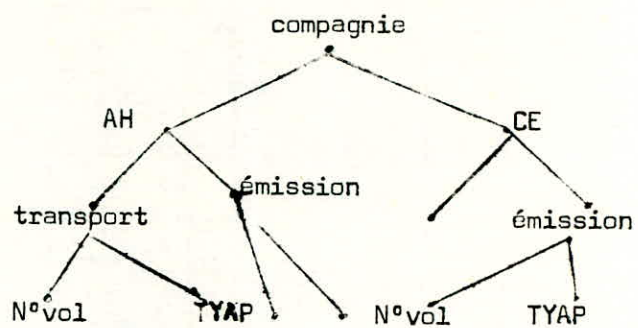


Prédicat numéro



Prédicat pool



Prédicat tempsPrédicat compagnie

3-4 Construction du Thésaurus

Le Thésaurus doit refléter les relations sémantiques entre les descripteurs de sortie et d'entrée des différents documents. Donc, pour pouvoir le construire on pourrait le faire à partir des relations.

Exemple:

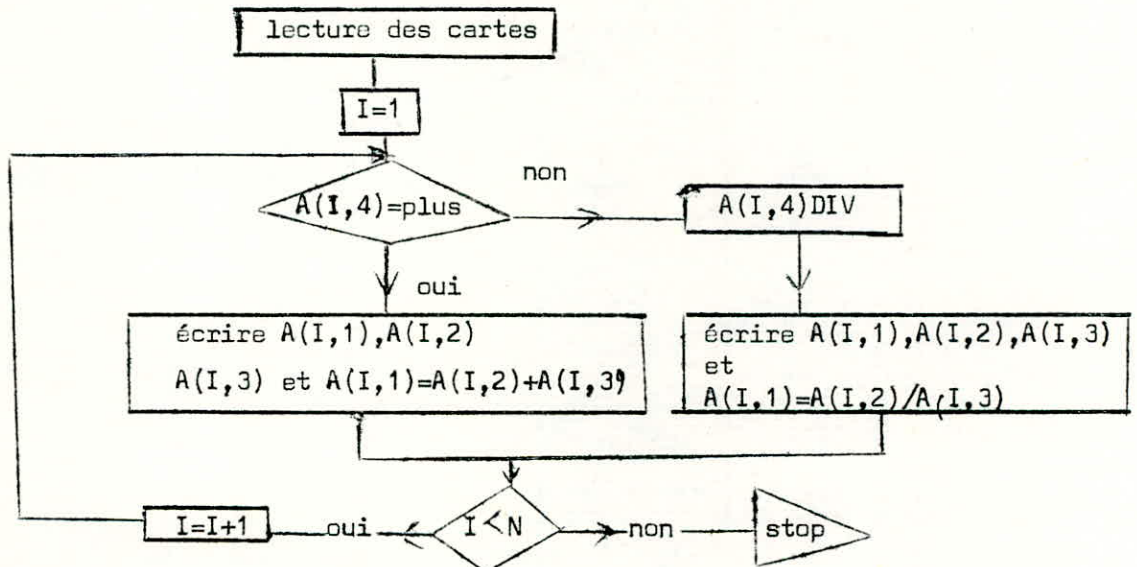
Si dans la carte, on perfore :

TNBP NBAH NBCE PLUS

On doit avoir en sortie:

TNBP NBAH NBCE TNBP = NBAH + NBCE

organigramme de construction du Thésaurus :



Le Thésaurus doit avoir la forme suivante :

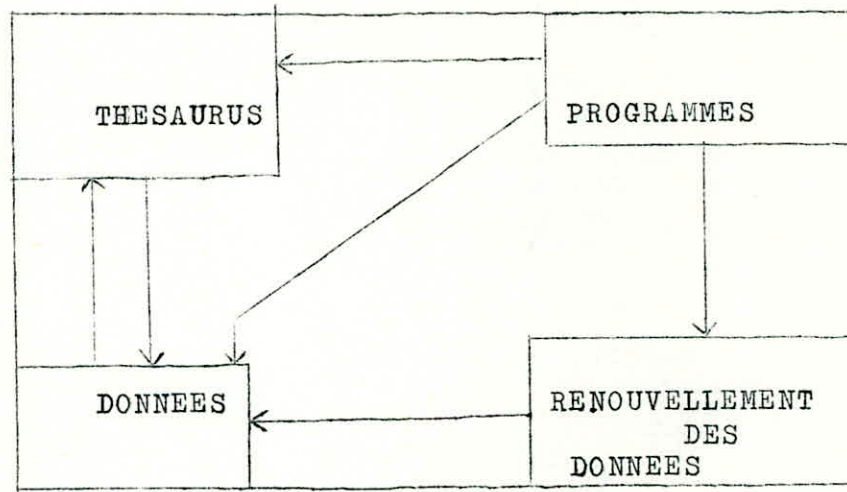
objet de sortie / objet d'entrée 1 / objet d'entrée 2 / ... / procédure de traitement

DS	DE 1	DE 2	DE 3	PROCEDURES
TNBP	NBAH	NBCE		TNBP = NBAH + NBCE
TREP	REAH	RECE		TREP = REAH + RECE
PMOY	TREP	TNBP		PMOY = TREP / TNBP
NBAH	CVPAH			NBAH = CVPAH
NBCE	CVPCE			NBCE = CVPCE
TOTGR	PGR	SER	BEBE	TOTGR = PGR+ SER+ BEBE
TOTAGF	CVF			TOTAGF = CVF
TOTAGTY	CVTY			TOTAGTY = CVTY
TOTAGTC	CVTC			TOTAGTC = CVTC
TOTC	TOTAG			TOTC = TOTAG
TOTCF	TOTAGF			TOTACF = TOTAGF
TOTCTY	TOTAGTY			TOTCTY = TOTAGTY
TOTCTC	TOTAGTC			TOTCTC = TOTAGTC
TOTGF	TOTCF			TOTGF = TOTCF
TOTGTY	TOTCTY			TOTGTY = TOTCTY
TOTGTC	TOTCTC			TOTGTC = TOTCTC
PAP	CVP			PAP = CVP
MCO	PCO	PR		MCO = PCO × PR
PAG	CVG			PAG = CVG
TGM	TMO			TGM = TMO
TRE	RECT			TRE = RECT
NBPAH	CVAH			NBPAH = CVAH
REPAH	REAH			REPAH = REAH
NBPCE	CVCE			NBPCE = CVCE
REPCE	RECE			REPCE = RECE
NBT	NBPAH	NBPCE		NBT = NBPAH + NBPCE
TOTMO	TOTJ			TOTMO = TOTJ
MMF	MMEV			MMF = MMEV × CP
STR	STR			STR = STR
DAT	DAT			DAT = DAT
DIST	DIST			DIST = DIST
COF	CHARG			COF = CHARG
FR	FR			FR = FR
PMPAP1	PMPAP1			PMPAP1 = PMPAP1
PMPAP2	PMPAP2			PMPAP2 = PMPAP2
PMFRET	PMFRET			PMFRET = PMFRET
PMPOST	PMPOST			PMPOST = PMPOST
SOFAP1	SOFAP1			SOFAP1 = SOFAP1
SOFAP2	SOFAP2			SOFAP2 = SOFAP2

SYSTEME SEMANTIQUE DE TRAITEMENT

4-1- CONFIGURATION GENERALE :

On peut représenter le système sémantique de traitement par la forme :



Dans le système on adonc le Thésaurus, les données et le renouvellement des données. Un système sémantique doit préparer les données par appel.

4-2- CONSTITUTION DES APPELS :

L'appel doit être formulé sous une forme standard et canonique :

<Appel> ::= <Objet>< Prédicat>/< Objet><Prédicat>
 < Prédicat > ... <Prédicat> /< Objet><Prédicat>
 ... <Prédicat> <Procédure>/<Descripteur>
 <Procédure >

<Procédure> ::= <Opérateur> /< Opérateur> ... <Opérateur >

<Objet > ::= <Expression > / <Terme >

<Expression> ::= <Terme> /< Terme > <Expression >

<Terme > ::= <Identificateur d'une notion >

<Identificateur d'une notion> ::= <mot qui a un sens dans le système >

<Prédicat> ::= <mot ou expression caractérisant les propriétés d'un objet >

L'appel doit donc décrire les informations demandées pour la gestion ou les moyens de sa réception à la base des données indiquées dans l'appel. Il est une composition de la description des données nécessaires ou comme un appel d'une procédure avec l'indication des paramètres à traiter.

Les termes sont les notions de base qui reflètent les composants de la structure du système de gestion c'est à dire les éléments (moyens techniques, moyens humains etc...) leurs relations (les liaisons, les fonctions etc...) et leurs caractéristiques (temporaires et spatiale).

Un appel est donc toujours une concaténation des termes :

$$A_k = \bigcap_i t_{ik}$$

Exemple :

Si on a à rechercher le nombre de passagers dans la relation ALGER - PARIS la formulation de l'appel se fera :

 TNBP ALG-PAR

On remarque que dans l'appel, l'objet (TNBP) est toujours un objet de sortie .

4-3 -ORGANISATION DES DONNES :

L'organisation des données se fait sous forme matricielle. En ligne on mettra les objets d'entrée associés aux objets de sortie et en colonne on aura les différents prédicats organisés sous forme hiérarchique. Ainsi avec les indications de la ligne et de la colonne, nous aurons l'information quantitative se rapportant au descripteur d'entrée associé.

θ \ P	P_1^1	$P_2^1 \dots P_p^1$	$P_1^2 \dots P_q^2$	$P_1^3 \dots P_d^3$	$P_i^4 \dots P_g^4$
θ_1	0	1.....0	0.....1	0.....0	0.....0
θ_2	1	1.....0	1.....0	0.....0	1.....0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
θ_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
θ_k	0.....00	0.....1	1.....0	1.....1

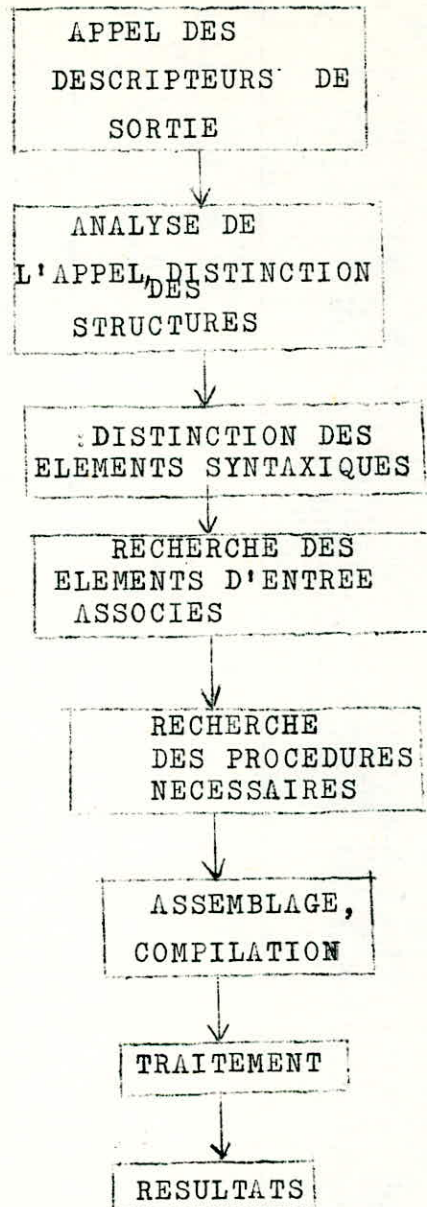
Exemple :

Si dans une ligne du thésaurus on a :

$$TNBP \quad NBAH \quad NBCE \quad TNBP = NBAH + NBCE$$

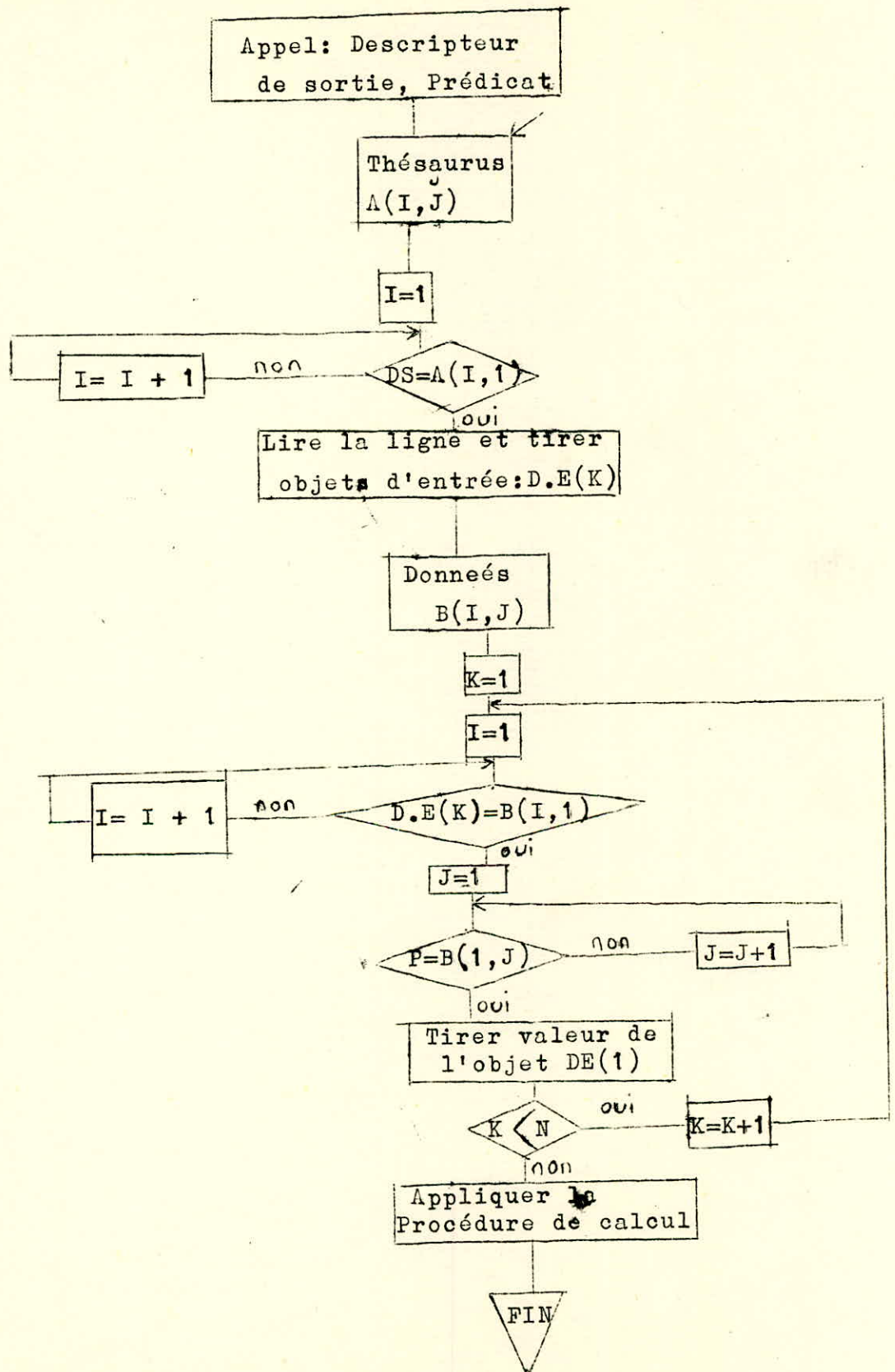
Dans la matrice des données on doit avoir :

θ \ P	ALG-PAR	ALG-LOND	JOUR	SEMAINE	MOIS
NBAH	100	150	250		
NBCE	50	100	150		



Un tel système fonctionne par appel, le système analyse l'appel recherche dans le thésaurus les objets d'entrée associés à l'objet de sortie figurant dans l'appel puis recherche leurs valeurs dans les données en fonction du ou des prédicats figurant dans l'appel et applique la procédure de calcul indiquée dans le thésaurus.

ORGANIGRAMME D'ANALYSE DE L'APPEL



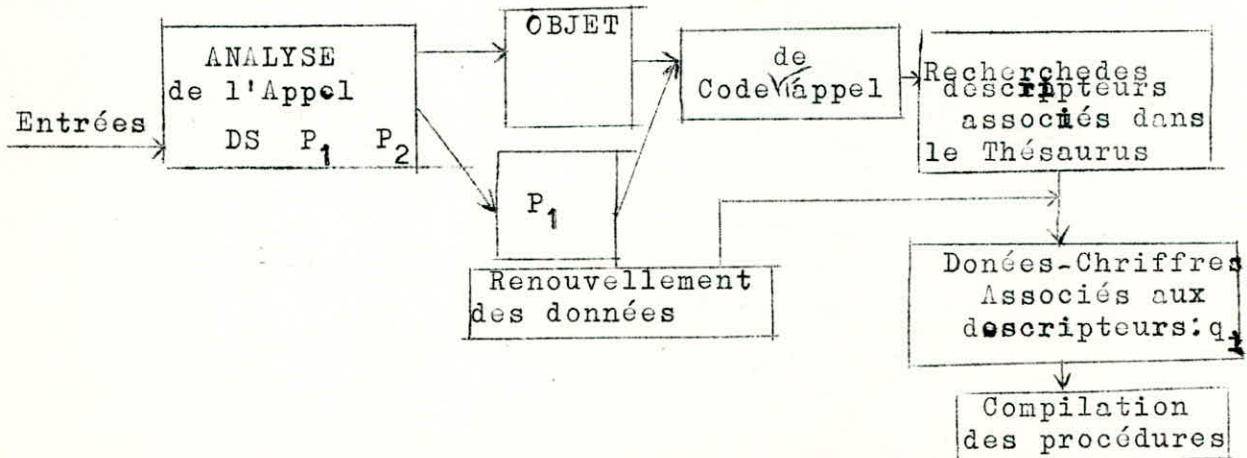
Avec l'analyse de l'appel, on aura les résultats désirés en fonction des valeurs des descripteurs figurant dans les données. On les données varient suivant les jours, il serait donc nécessaire d'établir un programme de renouvellement des données.

Les prédicats, dans la matrice des données, étant organisés sous forme hiérarchique, c'est à dire que les prédicats de niveau le plus bas sont inclus dans les prédicats de plus haut niveau, et ce pour chaque classe de prédicat. Ainsi, les jours sont inclus dans la semaine incluse dans le mois qui lui même est inclus dans l'année.

	J ₁	J ₂	J ₃	J ₄	J ₅	J ₆	J ₇	S ₁	J ₁	J ₂	J ₃	J ₄	J ₅	J ₆	J ₇	S ₂
θ ₁																
θ ₂																
⋮																
⋮																
θ _x																

Il faudrait donc, au fur et à mesure des jours remplir la matrice des données et faire le cumul au bout de la semaine, au fin de mois et de l'année.

D'une manière générale, le système peut être présenté de la façon suivante :



Un système avec un tel principe de traitement peut travailler en format libre et sans documents.

--- CHAPITRE V ---

Pour illustrer cette étude sur un système de traitement nous avons pris comme exemple le traitement partiel de la tâche n)4 " Transports Passagers " .

1) Thésaurus de cette tâche :

DS	DE1	DE2	RELATIONS
TNBP	NBAH	NBCE	TNBP=NBAH+NBCE
TREC	REAH	RECE	TREC=REAH + RECE
PMOY	TREC	TNBP	PMOY= TREC/TNBP

2) Matrice des données :

Dans ce tte matrice doivent figurer les objets d'entrée en ligne et en colonne, les prédicats classés sous forme hiérarchique .

Les données de cette matrice sont introduites au fur et à mesure des jours le programme renouvellement des données.

	J ₁	J ₂	J ₃	J ₄	J ₅	J ₆	J ₇	S ₂
NBAH								
REAH								
NBCE								
RECE								

3) CODIFICATION :

Pour faciliter la programmation, nous avons codifié le thésaurus et la matrice des données de façon à entrer simultanément dans les deux tableaux en donnant le même code aux colonnes du thésaurus et ligne de la matrice des données.

NBAH : 1
 NBCE : 3
 REAH : 2
 RECE : 4
 TREC : 2
 TNBP : 1
 + : 0
 / : 1

1	3	0
2	4	0
2	1	1

CODIFICATION DU THESAURUS :

4) Appel :

Si on veut connaître le nombre TOTAL de Passagers (TNBP) pour tous les jours de la semaine, l'appel se formulera de la façon suivante :

TNBP J₁ J₂ J₃ J₄ J₅ J₆ J₇ S₁

Il sera codifié de la façon suivante

1 1 2 3 4 5 6 7 8

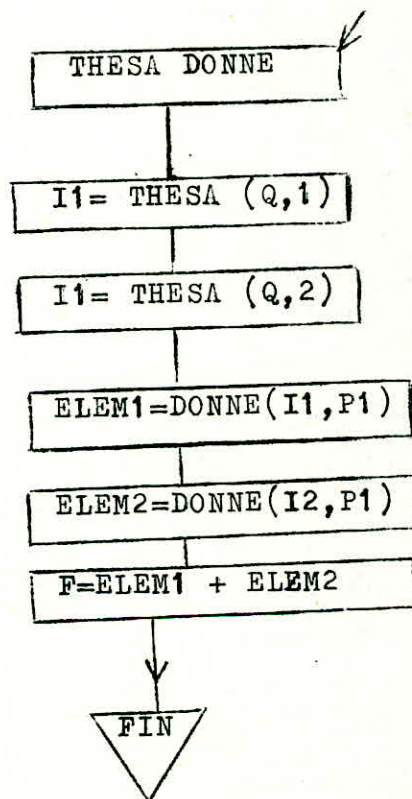
5) Traitement :

Le traitement de cet appel a été organisé comme suit:

- Un sous-programme qui permet d'entrer dans le thésaurus pour prendre les objets d'entrée associés à l'objet de sortie figurant dans l'appel, et dans la matrice des données pour repérer les valeurs associés aux objets d'entrée.

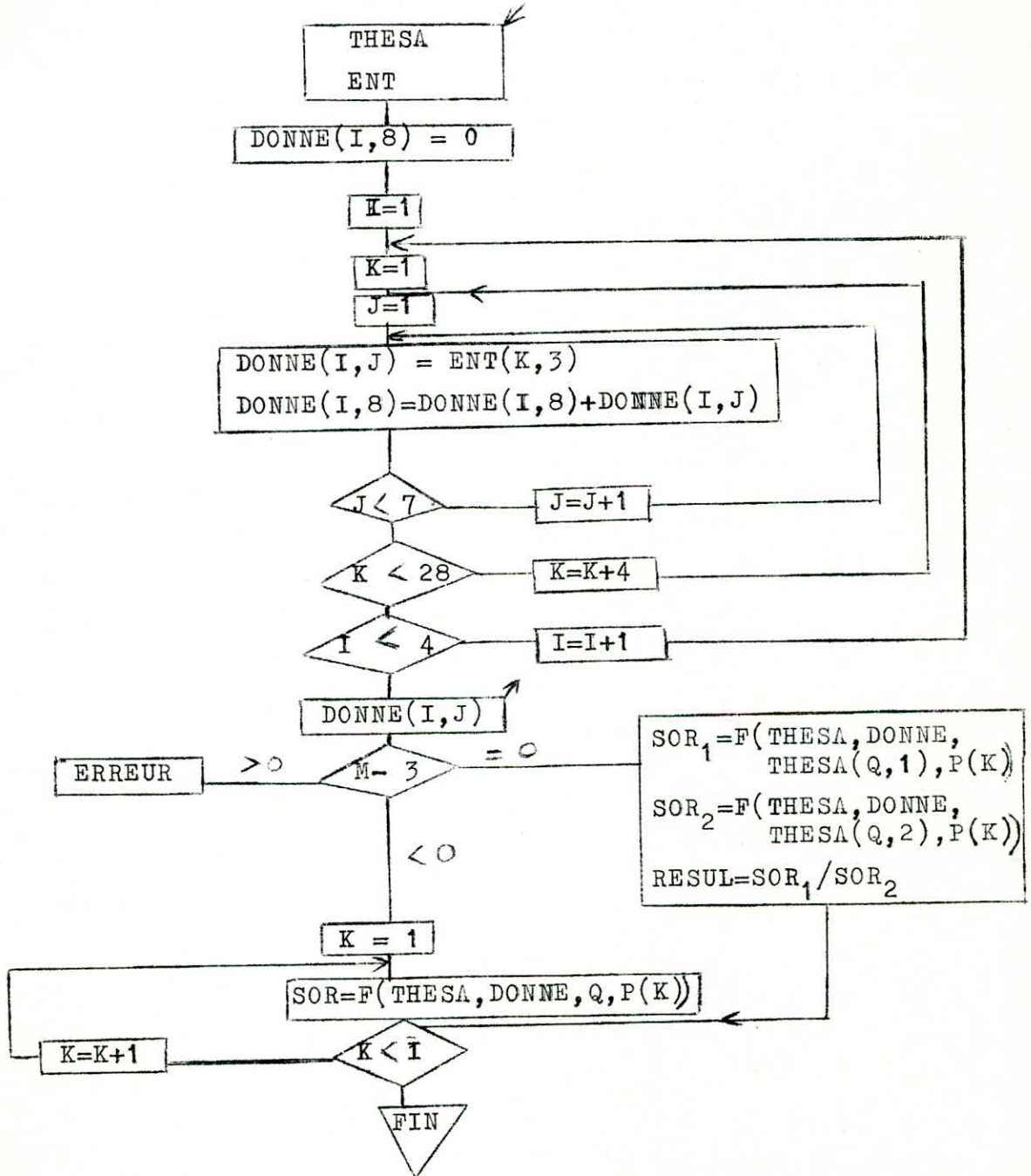
ORGANIGRAMMES :

- SOUS PROGRAMME FUNCTION F(THESA, DONNE, Q, P1)

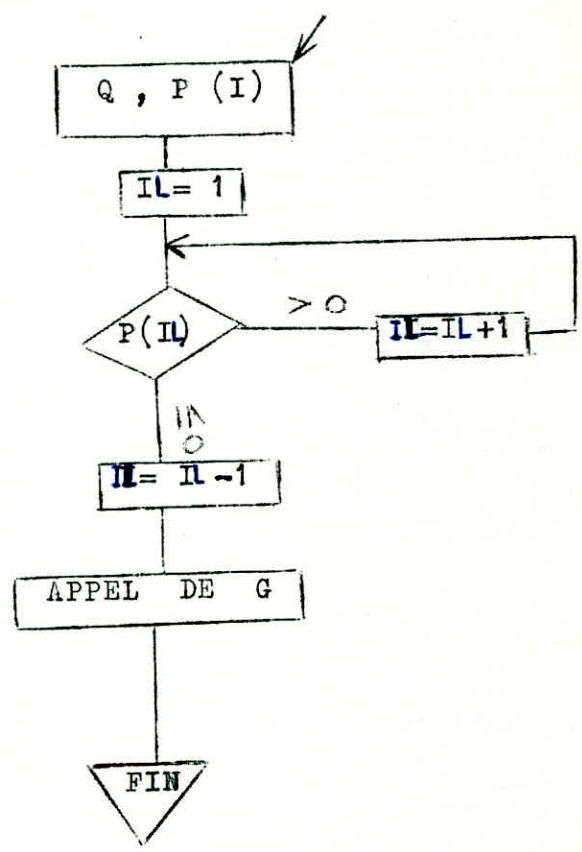


- UN sous-programme qui permet d'introduire les données dans la matrice des données et de calculer la valeur de l'objet de sortie.

6 Sous - Programme : SUBROUTINE G(Q,P,I)



- PROGRAMME PRINCIPAL :



Dans ce projet, on a exploité l'idée de l'analyse sémantique des documents pour la construction d'un système de traitement. Ainsi nous avons pu comparer les approches existants pour l'analyse du langage de l'information économique, défini l'approche sémantique de ce langage, établi sa morphologie et son algèbre partiellement, développé l'idée du Thésaurus en y incluant les procédures de traitement des descripteurs et enfin organisé les données sous forme matricielle ; ce qui nous a permis de diminuer la place occupée généralement par les données dans les autres systèmes ainsi que l'établi le renouvellement de ces données ce qui reflète le changement de l'état de l'objet de gestion. Les appels, présentés sous la forme canonique, sont analysés par le système qui travaille en format libre, ainsi on peut générer n'importe quel document utilisé par le système par une suite d'appels succesifs tous les descripteurs du document.

Cependant, cette étude n'est pas terminée, vu que les programmes ont été écrits au FORTRAN, langage non adapté au traitement des chaînes de caractères aussi l'exemple qu'on a traité est réduit par rapport aux résultats d'analyse bien que dans les documents que nous avons analysés, les procédures de traitement sont simples (addition, division), ceci est dû, en plus de la difficulté de programmation, au manque d'explication par l'entreprise concerné. Néanmoins, cette étude s'mantique de l'information offre un champ assez vaste d'application pour l'étude de système de gestion automatisé dans les entreprises car elle permet de définir une nouvelle méthode d'organisation des banques de données en liant les données avec leurs procédures de traitement.

-----BIBLIOGRAPHIE -----

- 1- M. ILIIN, etude des systemes economiques informationnels
ed. MESI, MOSCOU, 1968
- 2- SEMIOTIQUE ECONOMIQUE
Ed. "ECONOMIKA", MOSCOU, 1971
- 3- V. DOLIATOVSKI: cybernetique et gestion en economie
ed. IPK, ROSTOV, 1969
- 4- V. DOLIATVOSKI : Analyse et Synthèse des systèmes de traitement de
l'information
Ed. IRK, ROSTOV, 1972
- 5- MIDAY : Information RETRIEVAL
L. Willey And Sous, N-Y London 1970
- 6- R. NOSKOWITZ : Langage And Information
University of California, 1969
- 7- V. M. GLUCHKOV : Organisation et Traitement des Fichiers
KIEV, 1971
- 8- V. DOLIATOVSKI : N. Noveselkaia, E. Ovtchinnikov :
Etude et Construction d'un système intégré de gestion
de l'Institut de l'Enseignement Supérieur .IEA, Rostiv, 1972
- 9- R. C. GROSS, J. C. GARDIN, F. LEVY : AUTOMATISATION des Recherches
Documentaires - Un Modèle Général : Le SYNTOL.
