Contribution à la reconnaissance des structures syntaxiques en traduction mécanique

Alain Auroux

N° d'ordre :

THÈSE

présentée à

LA FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE GRENOBLE

pour obtenir

LE TITRE DE DOCTEUR DE TROISIÈME CYCLE

Mathématiques Appliquées

---

par

Alain AUROUX

Ingénieur I. R. G., I. M. A. G.

Licencié ès Sciences

CONTRIBUTION A LA RECONNAISSANCE DES STRUCTURES SYNTAXIQUES EN TRADUCTION MÉCANIQUE

Thèse soutenue le 23 juin 62 *devant la Commission d'examen :*

Monsieur J. FAVARD, Président

Messieurs J. KUNTZMANN, Examinateurs
B. VAUQUOIS
N. GASTINEL
R. GSELL

T H E S E

présentée à

LA FACULTE DES SCIENCES DE L'UNIVERSITE DE GRENOBLE

pour obtenir

le titre de Docteur de troisième cycle

Mathématiques Appliquées

Alain AUROUX

Ingénieur I.R.G.-I.M.A.G.

Licencié ès sciences

CONTRIBUTION A LA RECONNAISSANCE DES STRUCTURES SYNTAXIQUES EN TRADUCTION MECANIQUE

Thèse soutenue le 23 juin 62 devant la Commission d'examen

MM J. FAVARD — Président

J. KUNTZMANN
B. VAUQUOIS
N. GASTINEL
R. GSELL — Examinateurs

FACULTE DES SCIENCES DE L'UNIVERSITE DE GRENOBLE

Doyens honoraires :

M. FORTRAT P.
M. MORET L., Membre de l'Institut

Doyen :

M. WEIL L.

Professeurs :

MM. NEEL L. , Membre de l'Institut - Physique expérimentale
MOREL L. - Géologie et minéralogie
WOLFERS F. - Physique
DORIER A. - Zoologie
HEILMANN R. - Chimie organique
KRAVTCHENKO J. - Mécanique rationnelle
PARDE M. - Potamologie
BENOIT J. - Radioélectricité
CHENE M. - Chimie papetière
NOBECOURT P. - Micrographie papetière
BESSON J. - Chimie
WEIL L. - Thermodynamique
FELICI N. - Electrostatique
KUNTZMANN J. - Mathématiques appliquées
BARBIER R. - Géologie appliquée
SANTON L. - Mécanique des fluides
CHABAUTY C. - Calcul différentiel et intégral
OZENDA P. - Botanique
FALLOT M. - Physique industrielle
GOLVANI O. - Mathématiques
MOUSSA A. - Chimie nucléaire et radioactivité
TRAYNARD P. - Chimie générale
CRAYA A. - Hydrodynamique
SOUTIF M. - Physique générale
REEB G. - Statistiques mathématiques
REULOS R. - Théorie des champs
AYANT Y. - Physique approfondie
GALLISSOT F. - Mathématiques pures
Mlle LUTZ E. - Mathématiques générales
MM. BLAMBERT M. - Mathématiques
BOUCHEZ R. - Physique nucléaire
LLIBOUTRY L. - Géophysique
MICHEL R. - Géologie et minéralogie
BONNIER E. - Electrochimie

Professeurs sans chaire :

MM. SILBER R. - Mécanique des fluides
DESSAUX G. - Physiologie animale
MOUSSIEGT J. - Electronique
PILLET E. - Electrotechnique
BARBIER J.C. - Physique
BUYLE-BODIN M. - Electronique
PAUTHENET R. - Electrotechnique
Mme KOFLER L. - Botanique

Maîtres de Conférences :

MM. VAILLANT F. - Zoologie et hydrobiologie
DREYFUS B. - Thermodynamique
Mlle NAIM L. - Mathématiques
MM. PERRET R. - Servomécanisme
ARNAUD P. - Chimie
Mme BARBIER M.J. - Electrochimie
MM. BRISSONNEAU P. - Physique
COHEN J. - Physique
DEBELMAS J. - Géologie et minéralogie
Mme SOUTIF J. - Physique
MM. VAUQUOIS B. - Mathématiques appliquées
DEPASSEL R. - Mécanique des fluides
GERBER R. - Mathématiques
ROBERT A. - Chimie papetière
ANGLES d'AURIAC - Mécanique des fluides
BIAREZ - Mécanique physique
COUMES A. - Electronique
DODU J. - Mécanique des fluides
DUCROS P. - Minéralogie et cristallographie
GIDON P. - Géologie et minéralogie
GLENAT R. - Chimie
HACQUES G. - Calcul numérique
LANCIA R. - Physique automatique
PEBAY-PEROULA - Physique
GASTINEL - Chargé d'enseignement - Mathématiques appliquées
LACAZE A. - Chargé d'enseignement - Thermodynamique.

= =
=

Je tiens à exprimer ma profonde reconnaissance

à Monsieur le Professeur FAVARD qui a bien voulu me faire l'honneur de présider le Jury

à Monsieur le Professeur KUNTZMANN, Directeur du Laboratoire de Calcul de l'Université de Grenoble, qui a permis la réalisation de ce travail

à Monsieur VAUQUOIS, Maître de Conférences, auprès de qui j'ai trouvé les conseils et les indications qui m'ont permis de le mener à bien

à Monsieur GASTINEL, Maître de Conférences et à Monsieur GSELL, Chargé d'Enseignement à la Faculté des Lettres et Sciences humaines, qui ont bien voulu faire partie du Jury.

Je remercie également les membres du Centre d'Etudes pour la Traduction Automatique de Grenoble dont la collaboration me fut précieuse, Monsieur le Professeur CECCATO, Monsieur MARETTI et Mademoiselle ZONTA, de l'Université de Milan, qui m'ont donné les bases nécessaires à la réalisation d'une partie de ce travail.

I N T R O D U C T I O N

Sous sa forme la plus générale, le problème de traduction automatique se présente de la façon suivante :

Partant d'une langue source L_1 on veut arriver, par des moyens entièrement automatiques, à une langue cible L_2.

Une méthode très élémentaire consisterait à faire une traduction mot à mot : à chaque mot de L_1 on fait correspondre un mot de L_2 ; un tel procédé, s'il est simple, n'en est pas pour autant satisfaisant : en effet il ne permet pas de résoudre le problème des homographes, il risque de plus d'introduire des contre-sens en langue cible.

Une traduction correcte nécessitera donc l'étude syntaxique et aussi l'étude sémantique de la langue source. On est ainsi amené à procéder de la façon suivante :

On remplace la langue source L_1 par un langage formalisé L'_1, on cherche un algorithme A qui permette de passer de L'_1 à un langage L'_2 puis l'on passe de L'_2 à la langue cible L_2. L'_1 et L'_2 sont donc deux systèmes formels représentant les langages naturels L_1 et L_2.

L'analyse syntaxique automatique va consister à rechercher dans L'_1 des structures qui seront les images des structures syntaxiquement valables de L_1.

Une fois trouvées les structures syntaxiques de L'_1 il faudra faire un transfert de structure de L'_1 vers L'_2.

Aspect des recherches syntaxiques.

Il faudra faire une étude linguistique de L_1, dont le but est de trouver l'algorithme permettant de passer de L'_1 à L'_2, cet algorithme se fractionnant en analyse morphologique, analyse syntaxique, analyse sémantique et transfert de structures.

Préalablement à toute étude proprement syntaxique, il faudra définir le système L'_1, afin qu'il décrive le mieux possible le langage L_1, et qu'il permette la représentation d'un grand nombre de textes de L_1.

Une fois L'_1 défini et les matériaux linguistiques suffisants, il faudra trouver l'algorithme A permettant de passer de L'_1 à L'_2, ce qui, du point de vue syntaxique se traduira par des recherches de deux types.

a... Recherche d'une stratégie à adopter, cette stratégie devant permettre l'analyse syntaxique d'un texte quelconque de L_1.

b... Recherche de méthodes de programmation permettant l'accélération de l'analyse syntaxique automatique.

C H A P I T R E I

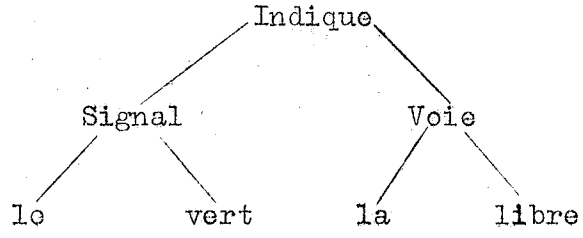
ROLE DE LA SYNTAXE ENTRE LA MORPHOLOGIE ET LA SEMANTIQUE

1-. Divers modèles d'analyse syntactico-sémantique.

Les modèles se répartissent en deux classes qui donnent toutes deux des descriptions de la phrase au moyen d'une structure d'arbre ; elles se distinguent de la façon suivante :

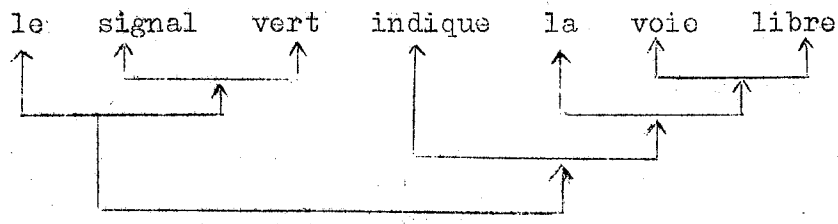
A-. La première place un mot de la phrase à chaque noeud de l'arbre, un même mot ne figurant qu'une fois dans le "stemma" (Tesnières, Hays, Lecerf) les différentes branches indiquent les liaisons de gouverneur à dépendant entre deux mots.

Exemple (Tesnières, éléments de syntaxe structurale, p.41)



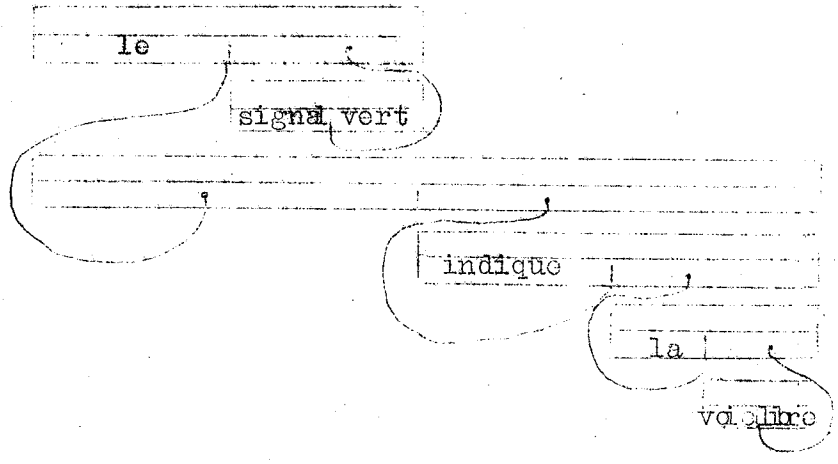
B-. La deuxième place les mots de la phrase au bout des branches terminales. et chaque noeud contient un nouvel élément fabriqué par le système (Chomsky).

Exemple



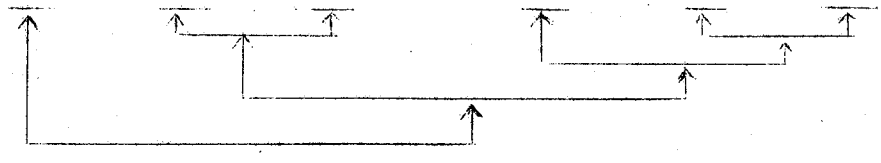
On peut ranger dans cette dernière classe le modèle des corrélogrammes de Ceccato, où chaque"rectangle" contient soit un élément fabriqué par le système, soit un mot :

Exemple



Nous nous attacherons dans ce qui suit aux modèles de la deuxième classe, et plus particulièrement à la première catégorie (bien qu'une étude technique du procédé des corrélogrammes soit donnée au chapitre 4)

Nous allons donc chercher la structure d'une phrase par composition binaire de mots ou de groupes de mots d'après des schémas du type :



L'algorithme permettant de trouver une telle structure sera fractionné en 3 parties :

- analyse morphologique
- analyse syntaxique
- analyse sémantique

2-. Rôle de l'analyse morphologique.

Le système formel L'_1 que l'on décrira à l'aide d'une méta-langue M devra être suffisamment complexe pour permettre l'étude d'un grand nombre de textes de L_1.

On peut définir l'analyse morphologique comme l'étude de chaque terme de L'_1 considéré individuellement. Autrement dit, l'analyse morphologique va faire correspondre à chaque forme de L_1 reconnue par L'_1 un terme qui sera l'élément de base des analyses syntaxiques et sémantiques (cf. Veillon, Veyrunes)

3-. Rôle de l'analyse syntaxique et de l'analyse sémantique : limite entre ces deux études.

L'analyse syntactico-sémantique est l'étude des combinaisons de termes (syntagmes) et l'élaboration des expressions de L'_1 correspondant à des structures dont on vient de voir la description schéma-

A-. L'analyse syntaxique va donner naissance à toutes les structures valables du point de vue des règles de construction formelle, sans faire appel à la signification des éléments qui entrent en jeu.

B-. L'analyse sémantique va avoir deux rôles :

a-. Faire un choix parmi les expressions trouvées grâce à l'analyse syntaxique (critère de validité sémantique).

b-. Remplacer les liaisons syntaxiques par des liaisons sémantiques, et lever les cas de polysémie (ce dernier point sera détaillé au chapitre II).

Il n'en reste pas moins que, si l'analyse morphologique peut être faite indépendemment, les analyses syntaxiques et sémantiques, bien qu'elles doivent être différenciées, resteront très proches, et que parfois certaines interférences pourront se produire (cf. chapitre 5).

4-. Nécessité d'une formalisation de la syntaxe.

Si dans les grammaires structurales classiques la limite entre la morphologie et la syntaxe-sémantique est assez bien discernable, il en va autrement pour celle existant entre syntaxe et sémantique qui n'apparaît pas clairement.

D'autre part les règles grammaticales énoncées font souvent appel au sens, implicitement au moins. Aussi pour utiliser ces grammaires en vue de la traduction mécanique, il faudra en élaborer une version purement formelle.

5-. Principales réalisations en matière d'analyse syntaxique.

Deux stratégies ont été envisagées jusqu'à présent :

A Stratégie dite de toutes les possibilités : elle consiste à rechercher toutes les structures acceptables. Elle a été adopté entre autres par Ceccato (théorie des corrélogrammes, cf. chapitre IV), par Lecerf (programme des conflits) etc...

B Stratégie de l'analyse prédictive, adoptée par Ida Rhodes, Oettinger, Kuno...

La première stratégie donne toutes les structures syntactiquement valables, mais est longue et encombrante car elle donne lieu à une arborescence rapide et nécessite souvent l'exploration de branches inutiles.

La seconde stratégie est plus rapide, moins encombrante, mais elle ne donne que la structure la plus probable, ce qui peut être parfois insuffisant ; de plus l'exploration de branches incohérentes ne peut être évitée, ce qui, avec cette stratégie, nécessite des retours en arrière.

La méthode d'analyse syntaxique que nous nous proposons d'adopter est dérivée de la première stratégie, mais nous allons étudier divers procédés permettant d'éviter au maximum le parcours de branches inutiles.

C H A P I T R E II

LA SYNTAXE DANS UN LANGAGE FORMALISE

1-. DESCRIPTION DU SYNTAGME ELEMENTAIRE

A Le point de départ de l'étude syntaxique sera le syntagme élémentaire : s.e.

Ce terme du système formel est défini de la façon suivante :

$$\langle s.e. \rangle \equiv \langle \upsilon \rangle \langle \text{n° d'unité lexicale} \rangle \langle Ku \rangle \langle C\Delta \rangle \langle Vgp \rangle \langle Vgc \rangle \langle Cs \rangle \langle C\sigma \rangle$$

υ : numéro séquentiel.
Ku : catégorie lexicale.
$C\Delta$: code dérivation.
Vgp : variables grammaticales permanentes.
Vgc : variable grammaticale contingente.
Cs : code syntaxique.
$C\sigma$: code sémantique.

Le code syntaxique sera lui-même formé d'un code dépendance et d'un code gouvernement :

$\langle Cs \rangle \Rightarrow \langle Csd \rangle \langle Csg \rangle$

B A toute forme ayant une représentation dans L'_1, on fera correspondre un certain nombre de syntagmes élémentaires, construits de la façon suivante :

a Chacun des syntagmes sera constitué par :

- un numéro séquentiel, (numéro progressif de la forme dans la phrase)
- un (ou plusieurs) numéros d'unité lexicale ,
- une (ou plusieurs) catégories lexicales, une catégorie lexicale groupant toutes les unités lexicales obéissant aux mêmes règles syntaxiques.
- des variables grammaticales permanentes ; ces variables sont fonction de Ku, et leur valeur est fixée par l'unité lexicale.
- des variables grammaticales contingentes, elles aussi fonction de Ku, mais leur valeur est fixée par la forme.
- un code syntaxique de dépendance.
- un code syntaxique de gouvernement.
- un code sémantique.
- un code de dérivation.

Toutes les variables grammaticales contingentes seront groupées en une seule, ce qui est toujours licite : en effet m variables pouvant prendre respectivement n_1, n_2,...n_m valeurs peuvent être remplacées par une seule variable, qui pourra alors prendre $n_1 \times n_2 \ldots \times n_m$ valeurs.

Exemple : soit une forme de L'_1 appartenant à la langue russe et appartenant à la catégorie lexicale "substantif" ; il y a alors 2 variables grammaticales permanentes :

- Le genre pouvant prendre 3 valeurs : masculin, féminin, neutre.
- L'animation pouvant prendre 2 valeurs : animé, inanimé.

La variable grammaticale contingente cas-nombre peut alors prendre 14 valeurs allant du nominatif singulier à l'instrumental pluriel.

L'unité lexicale U. L. attribue en général une et une seule valeur à chacune des variables grammaticales permanentes et la forme une ou plusieurs valeurs à la variable grammaticale contingente (souvent plusieurs dans les langues à flexion).

b un syntagme élémentaire est tel que chacune des variables prenne une et une seule valeur.

2-. ETUDE DE LA DEPENDANCE DES DIVERSES VARIABLES.

A Supposons d'abord les variables Vgp, Vgc, et Cs indépendantes entre elles, et considérons une forme à laquelle on a fait correspondre tous les syntagmes élémentaires que l'on peut en déduire. Divers cas peuvent alors se présenter:

a Tous ces syntagmes diffèrent seulement par la variable grammaticale contingente Vgc (c'est le cas, en français des substantifs invariables tels que "os") et peut-être par la valeur des codes syntaxiques. Il y a alors homographie interne.

b Ces syntagmes appartiennent à plusieurs unités lexicales correspondant à des catégories lexicales différentes, il y a alors homographie externe (à l'intérieur de laquelle il peut y avoir des homographies internes). Exemple : "forme", en français, est un mot appartenant à 2 unités lexicales auxquelles correspondent 2 catégories lexicales verbe et substantif ; en tant que verbe, "forme" présente une homographie interne.

c Ces syntagmes diffèrent uniquement par l'unité lexicale à laquelle ils appartiennent : du point de vue morphologique et syntaxique, ils sont alors identiques ; il y a alors polysémie.

Le mot français "glace" possède à la fois les 3 particularités précédentes : il appartient à 4 unités lexicales, l'une ayant comme catégorie lexicale la catégorie verbe, les 3 autres appartenant à la catégorie "substantif", d'où homographie externe ; les syntagmes correspondant à la catégorie verbe sont des homographes internes, et il y a polysémie entre les syntagmes appartenant à la catégorie substantif.

d. Ces syntagmes diffèrent par leurs unités lexicales, et par l'une au moins des variables Vgp, Vgc, Cs, mais ont même Ku ; suivant les cas on aura homographie externe ou polysémie c'est le cas du mot français "suis" auquel correspond 3 syntagmes élémentaires, dont deux, correspondant à la catégorie verbe, diffèrent seulement par leur code syntaxique; dans la phrase "je suis un chemin" il y a polysémie, alors que dans "tu suis un chemin" il y a seulement homographie externe.

L'analyse syntaxique doit permettre de résoudre les cas d'homographie (interne et externe) mais seule, l'analyse sémantique permettra de résoudre les polysémies.

B Il se peut que l'hypothèse d'indépendance des variables envisagée plus haut ne soit pas vérifiée ; par exemple, l'une des variables Vgp pourrait dépendre de Vgc, l'unité lexicale étant la même pour les syntagmes en question. Nous verrons ultérieurement l'incidence que cela peut avoir sur l'analyse syntaxique.

3-. DESCRIPTION D'UN ALGORITHME D'ETUDE SYNTAXIQUE.

A Nous appellerons fonction syntaxique, que l'on notera par φ, un opérateur qui s'appliquant à un couple de syntagmes élémentaires S_i, S_j, lui fait correspondre un élément S_k qui est de même nature que S_i et S_j et que l'on nommera "syntagme".

Supposons reconcées toutes les fonctions φ_p décrivant un phénomène linguistique et cherchons pour chacune d'elles la possibilité de liaison de deux syntagmes élémentaires, c'est-à-dire cherchons des relations du type $\varphi_p(S_i, S_j) = S_k$

Pour ce faire, nous devons dresser une liste exhaustive de tous les syntagmes élémentaires, abstration faite du numéro séquentiel, du numéro d'unité lexicale, de $C\Delta$ et de C_σ. Pour dresser cette liste,

considérons une catégorie lexicale $(Ku)_1$; cette catégorie fixe les variables grammaticales permanentes et contingentes ; considérons alors toutes les possibilités pour ces variables : nous obtenons une liste d'éléments ayant tous même catégorie lexicale, et où Vgp et Vgc prennent toutes les valeurs possibles. Remplaçons alors chacun des éléments de la liste ainsi constituée par des syntagmes élémentaires ayant même Ku, même Vgp, même Vgc et prenant toutes les valeurs possibles de Csd et Csg (ces codes syntaxiques sont déterminés par l'unité lexicale).

Par ce procédé, nous avons trouvé tous les syntagmes élémentaires correspondant à $(Ku)_1$; on répétant le processus décrit ci-dessus pour toutes les catégories Ku, on obtient bien une liste exhaustive des syntagmes élémentaires.

Pour déterminer toutes les liaisons du type $\varphi_p(S_i, S_j) = S_K$, il serait nessaire pour chaque fonction φ_p, d'étudier tous les couples S_i, S_j, de noter ceux qui donnent un résultat, ainsi que le S_K correspondant. Si n est le nombre de syntagmes élémentaires, m le nombre de fonctions φ , il faudrait faire n^2 x m études.

Il est alors évident que, supposant fait un tel travail, il serait possible, partant des syntagmes élémentaires correspondant aux divers mots d'une phrase de L_1 , de trouver toutes les liaisons pouvant exister entre les syntagmes correspondant à toutes les occurrences : il suffirait de prendre tous les couples S_i, S_j, avec $\mathcal{V}_{si} \neq \mathcal{V}_{sj}$, ($\mathcal{V}_{si}$ et $\mathcal{V}_{sj}$ étant les numéros séquentiels des mots auxquels correspondent S_i et S_j) de les comparer aux couples possibles précédemment trouvés, et de noter, lorsqu'il y a identité, la fonction φ_p liant S_i et S_j ainsi que le résultat S_K.

B Nous définirons un syntagme de la façon suivante :

<syntagme> = < υ > <Ku><Vgp > < Vgc ><Cs> <N> < F(φ_p, S_i, S_j)>

avec :

υ : numéro séquentiel.

Ku : catégorie lexicale.

Vgp, Vgc : variables grammaticales.

Cs : codes syntaxiques.

N : numéro de niveau.

F (φ_p, S_i, S_j) indications relatives à la formation de S_k.

Il est évident que, du point de vue syntaxique, la notion de syntagme est plus vaste que celle de syntagme élémentaire: un syntagme élémentaire est un syntagme. (en effet, $C\Delta$, C_σ et le numéro d'unité lexicale n'interviennent pas directement dans l'étude syntaxique).

Tous les éléments composant le syntagme S_k (à l'exception de υ et N qui sont en quelque sorte des numéros repères) sont calculés à partir de S_i et S_j par la fonction φ_p

Constituons alors une liste de tous les syntagmes : elle sera formée par celle précédemment décrite à laquelle on va ajouter tous les S_k, abstraction faite de υ, N et F(φ_p, S_i, S_j), à moins que ces syntagmes n'appartiennent déjà à la liste.

Pour terminer cette liste, il faudra lui ajouter les syntagmes obtenus à partir de ceux déjà écrits, à l'aide des fonctions φ_p ; le procédé de fabrication de cette liste est récursif, il permet de trouver tout les syntagmes de L'_1, ce nombre étant fini puisque le nombre de Ku, Vgp, Vgc et C_s l'est.

De même que la donnée de la liste des syntagmes élémentaires nous a permis de trouver toutes les liaisons existant entre les syntagmes élémentaires correspondant aux diverses occurrences d'une phrase à étudier, la donnée de la liste complète des syntagmes permet de trouver le syntagme correspondant à la phrase entière, et cela par au plus N-1 consultations de cette liste, N étant le nombre d'occurences de la

phrase . (Dans le cas d'ambiguité syntaxique et dans ce cas seulement, on obtiendra plusieurs syntagmes correspondant à une même phrase).

L'algorithme qui vient d'être décrit permet, théoriquement au moins, l'étude syntaxique d'une phrase quelconque de L_1, une fois donnés tous les syntagmes élémentaires correspondant aux occurences composant cette phrase.

C H A P I T R E III

STRATEGIE PAR COMPOSITION BINAIRE

1 Nous avons vu au chapitre II que la donnée de tous les syntagmes S et de toutes les fonctions φ conduisait à un algorithme d'étude syntaxique.

Cet algorithme serait d'une mise en oeuvre extrêmement simple sur un calculateur électronique : il se réduit pratiquement à une consultation de table, mais on se heurterait alors à deux difficultés majeures dûes à la taille de la liste exhaustive des syntagmes :
- cette table serait extrêmement longue à consulter-
- elle occuperait une place beaucoup trop grande dans la mémoire du calculateur-

Il faut donc, tout en conservant les principes mentionnés ci-dessus, trouver un autre algorithme, déduit du précédent, et qui permette un travail plus rapide et une meilleure occupation de la mémoire; il faut donc condenser les informations relatives aux syntagmes.

2 PROCEDE PERMETTANT L'ACCELERATION DE L'ANALYSE SYNTAXIQUE AUTOMATIQUE

A Considérons une famille d'homographes internes constituée de tous les syntagmes correspondant à une même occurrence ayant même catégorie lexicale.

a Supposons que tous les syntagmes constituant une famille d'homographes internes ne diffèrent entre eux que par la valeur que prend leur variable grammaticale contingente Vgc. Il est possible de regrouper tous ces syntagmes en un seul élément, que l'on notera S. E, obtenu en concaténant les valeurs prises par $\mathcal{U}$, Ku, le numéro d'unité lexicale, $C\Delta$, Vgp, Cs, $C\sigma$ et enfin l'union des valeurs prises par Vgc dans chacun des syntagmes de la famille d'homographes internes considérée.

L'algorithme de recherche des structures syntaxiques doit alors permettre de traiter les S. E. et non plus des syntagmes, puisque l'on

ne dispose plus de S_i et S_j individuellement, et que $(S.E.)_i$ correspond en fait à S_{i1}, S_{i2}... S_{im}.

Un exemple simple va permettre de voir l'intérêt du groupement en un S. E. d'une famille d'homographes internes :

Soit la fonction épithétique φ_p ; deux syntagmes élémentaires S_i et S_j seront tels que $\varphi_p(S_i, S_j) = S_k$ lorsque :

1- S_i a un Ku d'adjectif.
2- S_j a un Ku de substantif.
3- S_i et S_j ont même valeurs de variables grammaticales.
4- $I = \mathcal{V}_{sj} - \mathcal{V}_{si} - 1 = 0$ (intervalle nul entre les 2 occurrences auxquelles correspondent S_i et S_j)
5- Les codes syntaxiques de gouvernement de S_j sont compatibles avec les codes syntaxiques de dépendance de S_i.

Ces conditions étant réalisées, S_k aura les mêmes valeurs de variables grammaticales que S_j.

Il est évident que dans la liste de tous les syntagmes il y aura (si l'on suppose la condition 5 réalisée) 14 couples S_i, S_j pouvant être liés par la fonction φ, pour chaque valeur de V_{gp} et Cs.

Si le groupement par famille d'homographes internes a été fait, les règles relatives à la fonction φ_p vont alors être les suivantes :

1- $(SE)_i$ doit avoir un Ku d'adjectif.
2- $(SE)_j$ doit avoir un Ku de substantif.
3- $I = \mathcal{V}(SE)_j - \mathcal{V}(SE)_i - 1 = 0$
4- les codes syntaxiques de gouvernement de $(SE)_j$ doivent être compatibles avec les codes de dépendance de $(S\ E)_i$.
5- L'intersection des valeurs prises par les variables grammaticales contingentes de $(SE)_i$ et $(SE)_j$ doit être non vide, et les variables grammaticales permanentes doivent avoir mêmes valeurs.

Ces conditions étant réalisées, $(SE)_k$ aura mêmes valeurs de Vgp que $(SE)_j$, et pour valeur de Vgc l'intersection de celles de $(SE)_i$ et $(SE)_j$.

Dans la liste de SE il n'y aura plus qu'un couple $(SE)_i$ et $(SE)_j$ pouvant être lié par φ_p (toujours pour chaque valeur de Vgp et Cs) et un seul $(SE)_k$ résultant.

L'algorithme donne donc à la fois tous les syntagmes S_k, groupés en un $(SE)_k$ résultant de la liaison des S_i et des S_j par la fonction φ_p : en effet à chaque valeur non nulle de l'intersection des valeurs de variables grammaticales contingentes de $(SE)_i$ et $(SE)_j$ correspond un S_k.

Le groupement en un S.E des familles d'homographes internes nous a donc permis de réduire d'une façon importante la liste des syntagmes d'une part, et d'autre part de réduire le volume des informations correspondant à une occurrence du texte à traduire.

b ... Si maintenant on considère le cas d'homographes internes tels que les codes syntaxiques dépendent de la valeur prise par la variable grammaticale contingente, il est possible de grouper les syntagmes correspondants en plusieurs SE vérifiant l'hypothèse précédente, ou mieux en un seul SE à condition d'indiquer dans le code syntaxique lui-même la liaison existant entre Csg ou Csd et Vgc ; il sera alors nécessaire, si l'on adopte ce dernier groupement, de faire un calcul logique lors de l'étude de la compatibilité des codes syntaxiques des familles $(SE)_i$ et $(SE)_j$.

c ... Si l'une des variables grammaticales permanentes Vgp est fonction de la variable grammaticale contingente, il faudra comme dans le cas précédent soit écrire plusieurs familles d'homographes internes vérifiant l'hypothèse énoncée ci-dessus (paragraphe a), soit écrire un seul S.E. en indiquant dans Vgp le type de fonction liant Vgp à Vgc.

Dans ce dernier cas il faudra faire un calcul logique sur les

variables grammaticales lors de l'étude des possibilités de liaisons d'un tel S.E. avec une autre famille d'homographes internes.

B Supposons maintenant que les cas d'homographie externe aient été résolus, c'est-à-dire qu'à chaque occurrence du texte à traduire ne correspond qu'un seul S.E. ; nous parlerons ultérieurement de la légitimité de cette hypothèse en étudiant les possibilités de résolution des cas d'homographie externe.

Les règles de construction des syntagmes S_k à l'aide des fonctions φ peuvent se distinguer en deux groupes suivant le type de variables auxquelles elles font appel dans chacun des syntagmes S_i et S_j :

a Les règles qui se bornent à l'examen des valeurs prises par les catégories lexicales K et les numéros d'occurrence υ (ou plutôt les intervalles de ces occurrences, qui se rattachent aisément aux numéros par la relation :

$$I_{i,j} = \upsilon(SE)_j - \upsilon(SE)_i - 1$$

b Celles qui font appel aux variables grammaticales Vgp et Vgc , aux codes syntaxiques Csg et Csd.

Cherchons alors des règles du type $\varphi_q(Ku, Kv)$ permettant d'étudier globalement la possibilité de trouver une (ou plusieurs) fonctions φ liant les deux groupes d'homographes internes $(SE)_i$ et $(SE)_j$ ayant respectivement pour catégorie lexicale Ku et Kv.

Ces fonctions φ dépendent de Ku, Kv et, le cas échéant (lorsque $I \neq 0$), de la catégorie lexicale des SE séparant $(SE)_i$ et $(SE)_j$, c'est-à dire de ceux ayant un numéro υ_e tel que :

$$\upsilon(SE)_i < \upsilon_e < \upsilon(SE)_j.$$

Ces fonctions ψ ne sont pas des fonctionssyntaxiques : en effet elles donnent seulement une possibilité de liaison entre $(SE)_i$ et $(SE)_j$ et, lorsque la liaison existe effectivement, la catégorie lexicale Kw du résultat. Ce sont, en quelque sorte, des liaisons abstraites permettant de minimiser le nombre d'opérations inutiles qui serait très élevé dans le cas d'une recherche directe des fonctions φ_p. A une fonction ψ_q correspondra au moins une fonction φ_p, et, dans la grande majorité des cas, plusieurs de ces fonctions.

L'étude des possibilités de liaison de deux SE se fera donc en deux étapes :

-1 recherche d'une fonction $\psi_q(Ku, Kv) = Kw$

Si une telle fonction n'existe pas, il n'y a pas de liaison possible entre $(SE)_i$ et $(SE)_j$.

-2 Si une fonction de ce type existe, recherche des φ_p telles que : $\varphi_p [(SE)_i, (SE)_j] = (SE)_k$, sachant que si une telle liaison existe le groupe d'homographes internes $(SE)_k$ aura Ku pour catégorie lexicale.

Les numéros p des fonctions φ pouvant éventuellement lier $(SE)_i$ et $(SE)_j$ sont donnés par le numéro q de la fonction ψ considérée. En d'autres termes nous avons fractionné l'ensemble des fonctions en sous-ensembles tels qu'à l'intérieur de chacun d'eux, Ku, Kv, I soient identiques.

c En résumé nous avons remplacé l'algorithme permettant de décider s'il est possible de lier 2 syntagmes à l'aide d'une fonction φ_p par un algorithme permettant de décider si deux familles d'homographes internes peuvent être liées par cette même fonction φ_p, puis nous avons fractionné ce dernier, ce qui doit permettre un allègement et une accélération du procédé de reconnaissance des structures syntaxiques.

3 DETERMINATION DES DIVERS CONSTITUANTS D'UN SYNTAGME

Rappelons la définition du syntagme :

$$<\text{Syntagme}> \equiv <\upsilon><Ku><Vgp><Vgc><Cs><N><F(\varphi_p, S_i, S_j)>$$

A Les valeurs des variables Ku, Vgp, Vgc, Cs, sont déterminées, comme nous l'avons vu précédemment, par φ_p et les valeurs des variables correspondantes de S_i et S_j.

B N est un numéro de niveau ; pour un syntagme élémentaire, N = 0 ; pour un syntagme résultat de la liaison de 2 syntagmes élémentaires, N = 1. Dans le cas général, on a :

$$N = \text{Max}\ (N_1, N_2) + 1$$

N_1 et N_2 étant les numéros de niveau des syntagmes S_i et S_j.

C υ est un numéro progressif ;

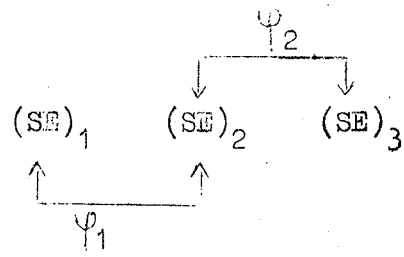
Pour un syntagme élémentaire υ est le numéro de l'occurrence dans la phrase. Nous verrons ultérieurement comment déterminer le numéro progressif d'un syntagme dans le cas général.

D $F(\varphi_p, S_i, S_j)$ représente un code permettant de décrire la façon dont a été obtenu le syntagme. Ce code comprendra les numéros progressifs de S_i et S_j ainsi que leurs numéros de niveau, et le numéro de la fonction φ_p :

$$<F(\varphi_p, S_i, S_j)> \equiv <\upsilon_{Si}><\upsilon_{Sj}><N_{Si}><N_{Sj}><p>$$

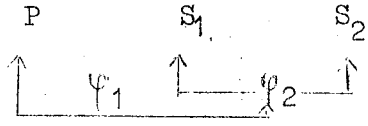
4 ETUDE DES PRIORITES DES FONCTIONS SYNTAXIQUES

Soient 3 familles d'homographes internes correspondant à 3 mots du texte à traduire $(SE)_1$, $(SE)_2$, $(SE)_3$ et supposons qu'il y ait liaison possible entre $(SE)_1$ et $(SE)_2$ d'une part, entre $(SE)_2$ et $(SE)_3$ d'autre part, et ceci à l'aide des fonctions φ_1 et φ_2.

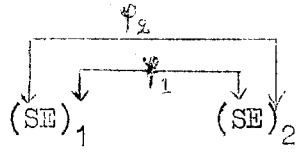


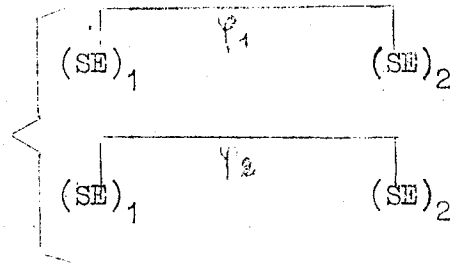
L'une de ces liaisons devra être faite avant l'autre, ceci étant déterminé par les priorités relatives des fonctions φ_1 et φ_2 : la liaison faite la première (c'est-à-dire celle de niveau le plus bas) sera celle correspondant à la fonction syntaxique la plus prioritaire.

Soit par exemple la suite préposition, P, substantif gouverné par la préposition, S_1, substantif au génitif, S_2, gouverné par S_1 ; il existe une fonction φ_1 liant P à S_1, mais elle est moins prioritaire que celle liant S_1 à S_2 ; si φ_2 est cette dernière, on aura donc le résultat suivant :



Dans le cas où il peut exister plusieurs fonctions syntaxiques liant deux familles d'homographes internes, il faudra alors faire chacune des liaisons, c'est-à-dire poursuivre dans la suite plusieurs études distinctes :





Il faut bien remarquer que deux syntagmes ne peuvent être liés que par une seule fonction syntaxique, alors qu'il peut en exister plusieurs liant deux groupes d'homographes internes.

C H A P I T R E IV

REALISATION SUR CALCULATEUR DE L'ALGORITHME D'ETUDE SYNTACTICO-SEMANTIQUE DU GROUPE DE MILAN (ECOLE OPERATIONNELLE ITALIENNE)

1-. Principes de la méthode.

A Corrélogrammes- Toute liaison entre deux mots ou groupes de mots d'un langage L_1 est une corrélation qui comprend toujours 3 "éléments" :

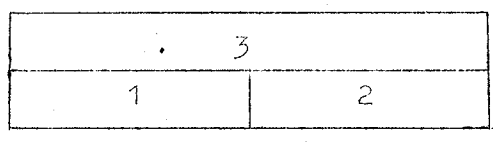
- un premier corrélé
- un deuxième corrélé
- un corrélateur

Le corrélateur pouvant être implicite ou explicite.

Soit par exemple la phrase "Pierre et Paul arrivent".

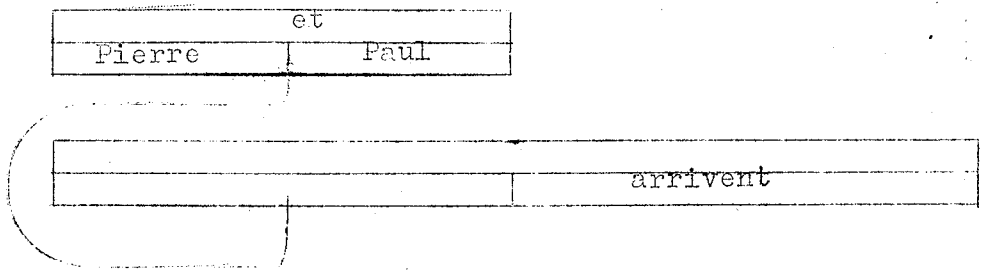
Une première corrélation existe entre "Pierre" et "Paul", ces mots étant respectivement premier et deuxième corrélés, cette corrélation est marquée par "et" corrélateur explicite. Il y a une deuxième corrélation entre "Pierre et Paul" et "arrivent" le corrélateur étant implici

B Une corrélation sera représentée par un rectangle composé de 3 cellules, contenant respectivement le 1er corrélé (1), le 2e corrélé (2), le corrélateur (3).



Dans chacune des cellules on peut mettre soit un mot soit un rectangle entier, qui sera alors plus petit que celui dans lequel il rentre comme premier, deuxième corrélé ou corrélateur.

Ainsi la phrase ci-dessus sera représentée de la façon suivante :



2-. Relation entre les corrélogrammes et les syntagmes, entre les corrélations et les fonctions syntaxiques.

A Les éléments de départ de l'analyse syntactico-sémantique sont les "matrices mots" dont chaque ligne représente en fait une famille d'homographes internes ; dans le cas où les variables grammaticales permanentes, les codes syntaxiques ou sémantiques dépendent des variables grammaticales contingentes, on crée autant de lignes de "matrices- mots" qu'il est nécessaire pour rendre les variables indépendantes entre elles. (cf. chapitre II P. 3) A chaque ligne de "matrice-mot", on associera 3 ou 2 rectangles élémentaires, suivant que, pour cette ligne, le mot peut être corrélateur ou seulement corrélé. Cela découle des 2 postulats suivants :

1 Tout mot qui n'est pas corrélateur peut-être premier ou deuxième corrélé.

2 Tout corrélateur explicite peut être premier ou deuxième corrélé (cela afin de permettre l'étude de phrases telles que "et et ou sont des corrélateurs").

B Chaque corrélateur explicite est caractérisé par un indice I_c supérieur ou égal à 1, l'indice $I_c=0$ étant réservé aux corrélations implicites.

Chaque rectangle saturé(c'est-à-dire dans lequel les 3 cellules élémentaires sont remplies) est caractérisé par un indice de corrélation I_{cR} et, éventuellement, par un sous-type.

Il est à remarquer que l'ensemble I_{cR}, sous-type correspond aux fonctions syntaxiques φ_p dont nous avons précédemment parlé, alors que l'indice de corrélation I_{cR} seul correspond, dans une certaine mesure aux pseudo-fonctions syntaxiques ψ_q.

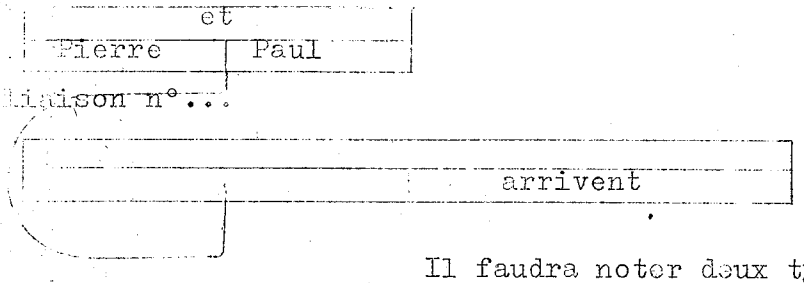
3-. De même qu'à un mot est associé une matrice-mot ayant plusieurs lignes et un certain nombre de rectangles, à un groupe de mots analysé ou à une phrase sera associé une matrice-produit et un corrélogramme.

A La matrice-produit contient les informations syntaxiques et sémantiques: il y a entre matrice-produit et matrice-mot la même analogie qu'entre un syntagme et un syntagme élémentaire.

B Le corrélogramme contient les informations relatives aux I_c, aux I_{cR} et sous-types ; il définit la position des mots dans les rectangles et des rectangles entre eux.

La position relative de deux rectangles fait ressortir le plus grand rectangle, c'est-à-dire celui qui joue le rôle prépondérant, ainsi que les premier et deuxième corrélés de ce rectangle.

Si l'on reprend l'exemple précédent : "Pierre et Paul arrivent" le corrélogramme est le suivant:



I_{cR} ≠ Ss-type =...

I_{cR} / Ss-type =...

Il faudra noter deux types d'indications :

a ... Indications relatives à chacun des rectangles, c'est-à-dire ici :

1er rectangle : 1er corrélé : "Pierre"
2e corrélé : "Paul"
Corrélateur : "et"
Fonction de corrélation (I_{cR} et Sous-type)

2e rectangle : 1er corrélé : 1er rectangle
2e corrélé : "arrivent"
Corrélateur : implicite
Fonction de corrélation (I_{cR} et sous-type)

b ... Indications relatives à la liaison :

- numéro des 2 rectangles liés (ici 1 et 2)
- type de liaison (c'est-à-dire position et taille relatives des 2 rectangles)

c ... La méthode d'analyse est basée sur une stratégie étudiant toutes les possibilités de liaisons entre rectangles et corrélogrammes, cette analyse se faisant mot après mot du début à la fin de phrase ; ainsi lorsque l'on a trouvé tous les corrélogrammes existant avec les n premiers mots, on va étudier les liaisons entre, d'une part ces corrélogrammes, d'autre part les rectangles correspondant aux (n + 1)eme mot, puis celles entre les corrélogrammes correspondant aux premiers mots et ceux que l'on vient de déterminer, et ainsi de suite.

4-. Divers mode d'obtention de corrélogrammes.

A Règles déduites des seuls corrélogrammes.

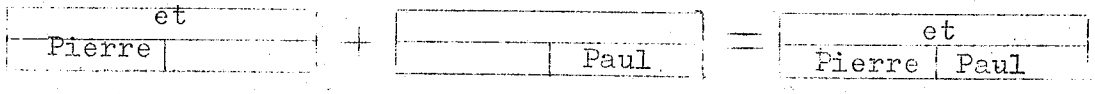
Nous dirons qu'un rectangle est saturé si les trois cellules le composant sont remplies (le corrélateur pouvant être implicite) et qu'il est libre s'il a au moins une cellule de corrélé non remplie.

Nous dirons qu'un corrélogramme est saturé si tous les rectangles qui le constituent le sont et qu'il est libre si son plus grand rectangle a une cellule de corrélé non remplie.

Il est évident que, quelque soit le corrélogramme, tous les rectangles qui le constituent, sauf peut-être le plus grand dans le cas d'un corrélogramme libre, sont saturés.

a. Insertion Il y aura possibilité de liaison entre deux corrélogrammes libres si l'un a une cellule premier corrélé libre et l'autre une cellule deuxième corrélé libre.

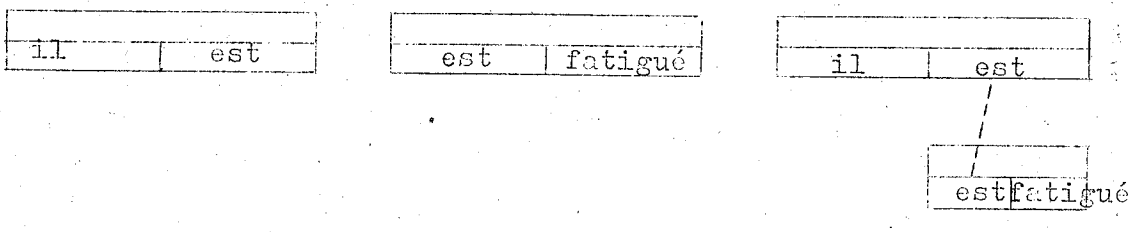
Exemple



Dans le cas d'une insertion le nombre de rectangles du corrélogramme résultat est égal à la somme des nombres des deux corrélogrammes donnés moins un.

b. Superposition Il y aura liaison entre deux corrélogrammes saturés lorsque le premier rectangle de l'un et le dernier rectangle de l'autre ont le même mot comme corrélé en position non homologue.

Exemple



Le type de liaison créé dans ce cas est fonction des indices corrélationnels des deux réseaux de départ ; pour qu'il n'y ait pas

d'impossibilité géographique, il faut que le rectangle faisant la liaison qui devient le plus petit des deux, soit le plus grand dans son corrélogramme origine.

B Règles déduites des matrices-mots (ou matrices-produit)

a Insertion - Il faut qu'il y ait accord entre les matrices-produit (ou mot) des deux corrélogrammes pour que la liaison soit possible. Cet accord porte d'abord sur les indices I_c qui doivent être les mêmes, puis sur les différentes variables constituant la matrice-produit : catégories lexicales, variables grammaticales, syntaxiques et sémantiques, et enfin sur l'ordre des mots constituant les corrélogrammes et leurs intervalles.

La recherche de cette possibilité d'accord constitue le contrôle ; il est réalisé à l'aide de "matrices-contrôle " qui sont des codes exprimant les règles d'accord.

Il y a autant de matrices-contrôle que d'indices I_c. Chaque matrice - contrôle est constituée par un certain nombre de lignes correspondant chacune à un $I_{c\ R}$ et sous type (c'est à dire à une fonction syntaxico-sémantique).

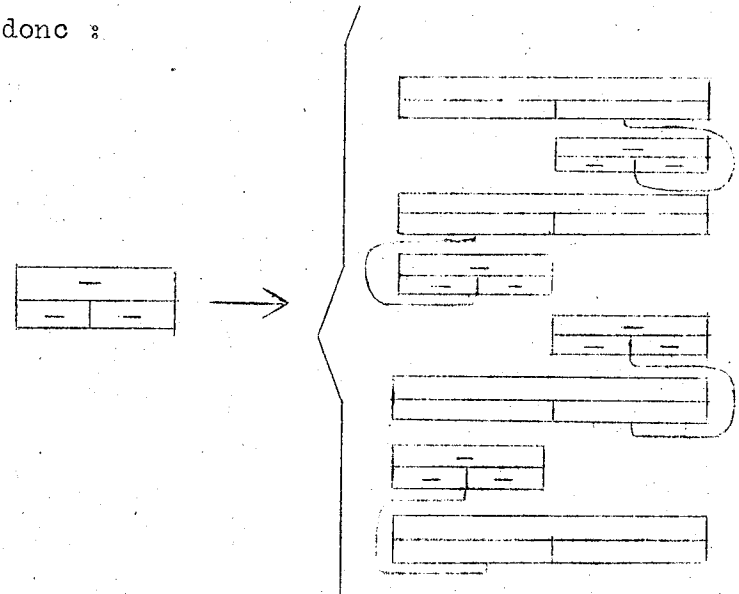
b Superposition - Les seules règles de la superposition ont déjà été énoncées ; il n'y a aucun contrôle dans ce cas.

c Un corrélogramme ne doit pas contenir plus de 2 fois le même mot et il le peut seulement dans le cas où ce mot sert de lien à une superposition.

Le résultat d'une insertion ou d'une superposition est toujours un corrélogramme saturé ; pour que, dans la suite, le procédé puisse continuer, il faut, par un artifice, déduire des corrélogrammes libres de ceux qui sont saturés. Cette opération s'appelle la reclassification ;

chaque corrélogramme saturé est en effet "reclassé" en 4 corrélogrammes d'indice de corrélation $I_c = 0$ en liant des 4 façons possibles le plus grand rectangle du corrélogramme origine à un rectangle vide ;

On a donc :



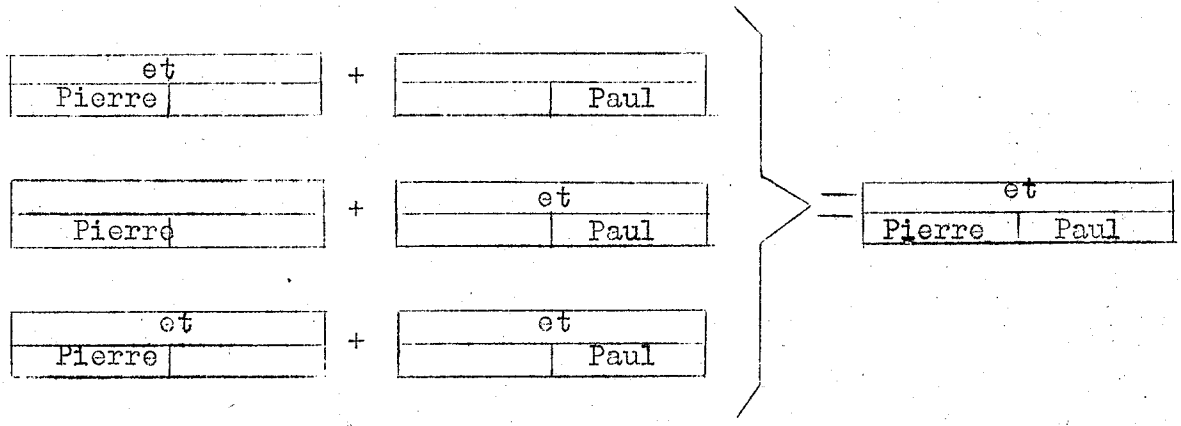
5-. Réduction du nombre des corrélogrammes.

Deux règles permettent de réduire légèrement le nombre des corrélogrammes :

1- Un corrélogramme qui a servi dans une superposition peut être éliminé.

2- Un corrélogramme qui a servi comme 2e corrélé dans une insertion peut être éliminé

Il est bien évident que, de plus, il faut vérifier, à chaque corrélogramme que l'on trouve, que ce corrélogramme est nouveau ; en effet le procédé est théoriquement ternaire (binaire seulement dans le cas où la corrélation n'a pas de marquant explicite) alors que la réalisation machine est binaire ; on est donc conduit à étudier, dans le cas de la phrase "Pierre et Paul arrivent" 3 liaisons qui donnent le même résultat :



Voilà ainsi résumé la méthode du groupe de Milan ; pour un exposé plus détaillé, voir "linguistic analysis and programming for mechanical translation".

Envisageons maintenant la programmation adaptée à une telle méthode.

6-. Programme réalisé sur calculateur I. B. M. 704 -7090.

A Organigramme général.

Le programme proprement dit a été divisé en un certain nombre de programmes plus petits, ces programmes communiquant entre eux par par un nombre restreint de points d'entrée et de sortie:

I- Entrée de la phrase à analyser.
II- Consultation du dictionnaire.
III- Fabrication des rectangles correspondant à un mot.
IV- Etude des possibilités de liaison d'après d'après la saturation des réseaux.
V- Insertion - Ce programme comprend 2 parties :
V-1 Contrôle.
V-2 Fabrication des réseaux.

VI- Superposition- Là encore deux parties :

VI-1 Etude des possibilités de superposition.

VI-2 Fabrication des réseaux.

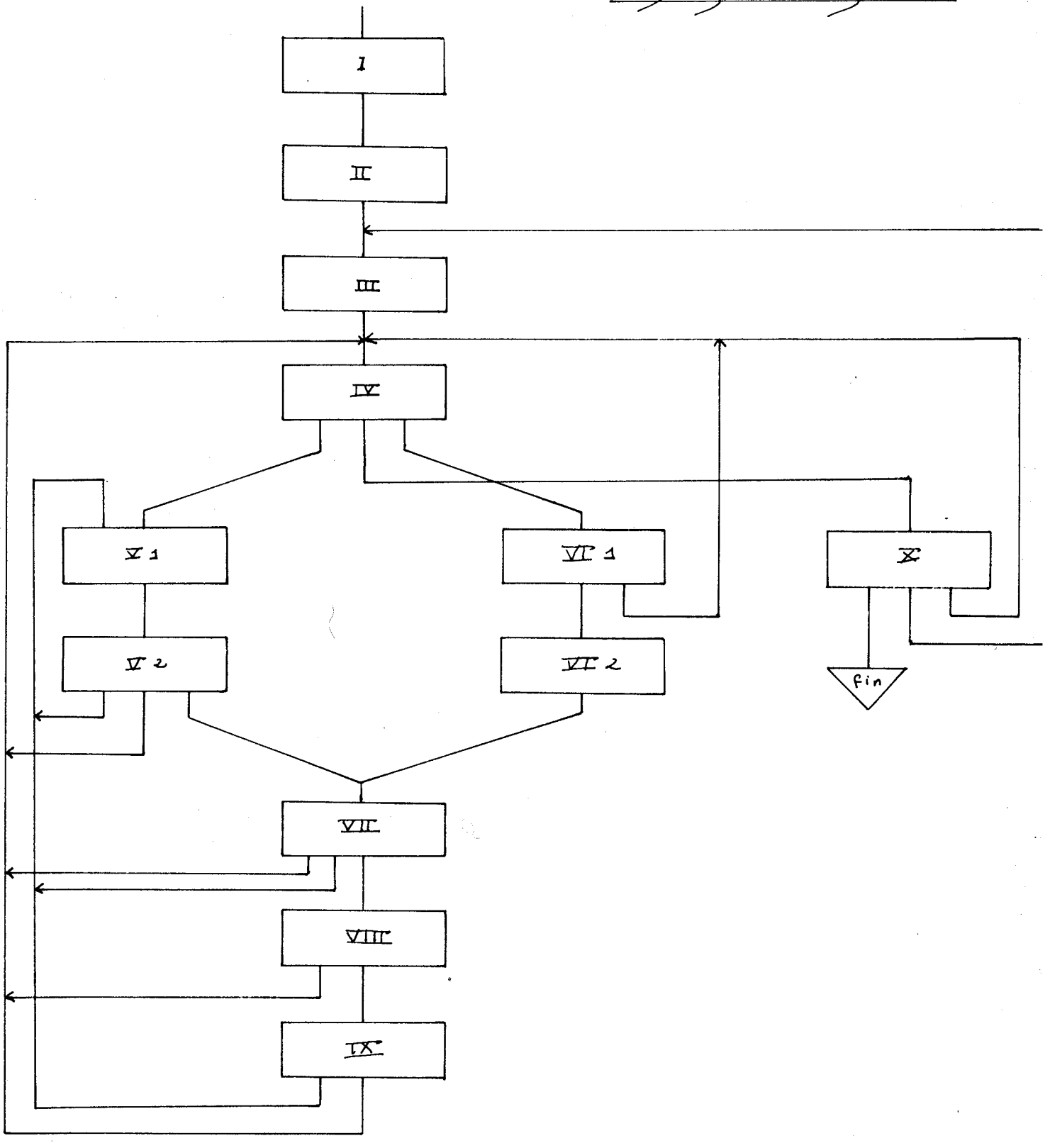
VII- Vérification que le réseau est nouveau.

VIII- Elaboration de la matrice-produit.

IX- Reclassification des réseaux saturés.

X- Mise à jour de la liste des corrélogrammes, recirculation.

Organigramme général-



B Organisation des informations dans la mémoire du calculateur.

a. Langage utilisé- Le programme a été écrit en SAP absolu, langage symbolique employé sur ordinateur IBM 704 ; puis adapté en FAP (en modifiant les ordres d'entrée et de sortie) afin de pouvoir être utilisé sur 7090.

On a utilisé l'adressage absolu des données (en effet il n'y avait pas sur 704 de programme moniteur, il était donc peu intéressant d'écrire un programme avec adressage relatif.)

b. Contraintes technologiques- L'ordinateur utilisé doit avoir une mémoire de 32.768 mots. Il y a de plus 3 bandes données et une bande résultats.

c. Organisation de la mémoire- Les 100 premiers mots sont laissés libres pour les programmes de servitude (chargement, dump).

Les mémoires 100 à 9.999 contiendront d'abord la suite des mots de la phrase à analyser, puis les réseaux corrélationnels (corrélogrammes).

Les mémoires 10.000 à 19.999 contiendront les matrices-mots et les matrices-produits.

Les mémoires 20.000 à 27.999 contiendront les données et tables (matrices-contrôles, tables diverses, masques).

Le programme sera rangé à partir de la mémoire 28.000.

d. Contraintes dues au programme- L'étude se fait phrase à phrase ; une phrase ne doit pas contenir plus de 127 mots ; chaque matrice-mot ne doit pas avoir plus de 31 lignes. De plus il ne faut pas que les corrélogrammes utilisent plus de 9.900 mémoires.

C Représentation des corrélogrammes.

a Codification. Chaque corrélogramme portera un numéro progressif k ; $(1 \leqslant k < 2^{12})$. Il n'y aura pas de renumérotation des corrélogrammes après la recirculation, c'est-à-dire que si l'un d'entre eux est effacé, il n'y aura pas de corrélogramme portant son numéro.

Soit α le nombre de rectangles d'un corrélogramme ; il faudra 3α mémoires pour décrire l'un d'entre eux.

Tous les corrélogrammes contiendront dans une première mémoire leur numéro progressif, ainsi que ceux des 1er et 2e corrélés. Nous désignerons par A ce type de mémoire.

On trouvera en plus les 3 types de mémoire qui suivent :

B contient les indications relatives à une liaison, c'est-à-dire les numéros des rectangles liés et le type de liaison (ou fonction de disposition).

C contient, pour un rectangle, son indice de saturation, son indice corrélationnel (I_c ou I_{cR} et sous-type suivant la saturation) et un numéro de fonction de saturation (cette indication fait double emploi avec l'indice de saturation).

D contient, pour chaque rectangle, les numéros d'occurrence et numéros de ligne de matrice-mot de chacun des éléments du rectangle.

En plus la 2e mémoire contient le nombre α de rectangles du corrélogramme, l'indice de saturation du corrélogramme, un indice d'effacement (dont la valeur est 1 si le corrélogramme est destiné à être effacé, 0 dans le cas contraire) et, si $\alpha \neq 1$, un indice δ $\delta = 1$ si le corrélogramme a été obtenu par superposition, 0 dans l'autre cas) (mémoires E et F). (Cf page 25).

Les 4 corrélogrammes déduits par reclassification d'un réseau corrélationnel saturé porteront le même numéro que ce dernier, et aucune indication en numéros de 1er et 2e corrélés ; on peut donc en appelant k_1, k_2, k_3 respectivement les numéros de corrélogrammes, de réseau 1er corrélé et de réseau 2e corrélé avoir les cas suivants :

- $k_1 = 0$ (alors $k_2 = k_3 = 0$) le corrélogramme correspond à un mot seul.
- $k_1 \neq 0$ $k_2 = k_3 = 0$ si $\alpha = 1$, le corrélogramme est issu de 2 mots.
  $\alpha \neq 1$, le corrélogramme est obtenu par reclassification.

$k_2 \neq 0, k_3 = 0$
$k_2 = 0, k_3 \neq 0$
} rien de spécial à signaler

$k_2 \neq 0, k_3 \neq 0$ Si le corrélogramme a été obtenu par superposition, k_2 et k_3 sont respectivement les numéros de réseaux récepteur et reçu.

b ... Organisation des listes de corrélogrammes.

Nous aurons toujours des comparaisons de corrélogrammes deux à deux. On appèlera liste B la liste des b-corrélogrammes premier opérande, liste A celles des adeuxième opérande et C celle des c résultats.

Le nombre de rectangles de chaque corrélogramme fixe l'encombrement de ce corrélogramme. Connaissant a, b, c, l'adresse de début de la liste B, on dispose de tous les éléments permettant de trouver l'adresse d'un quelconque corrélogramme de l'une des listes A, B, C, pourvu que l'on ait un numéro progressif.

c ... Evolution de la liste des corrélogrammes : recirculation.

Supposons la liste B constituée de tous les corrélogrammes correspondant aux n premièrs mots de la phrase à analyser, on rangera

alors dans la liste A les rectangles correspondant au mot n + 1 ; on constituera ensuite la liste C (par superposition ou insertion) ceci étant fait, par exploration de tous les corrélogrammes on constituera une nouvelle liste B contenant tous ceux restant des anciennes listes A et B, et une nouvelle liste A qui n'est autre que l'ancienne liste C.

Le procédé de recirculation est récursif, il s'arrêtera lorsqu'on ne trouvera plus de corrélogramme dans la liste C.

Au cours de l'exploration de la liste, on sera amené, lorsque certains corrélogrammes doivent être effacés, à condenser cette liste en remontant certains de ses éléments.

L'organigramme de la recirculation est alors le suivant, dans lequel on aura les paramètres :

c nombre de réseaux de la liste C

a " " " A Origine

b " " " B " "

a_1 " " " A Résultat

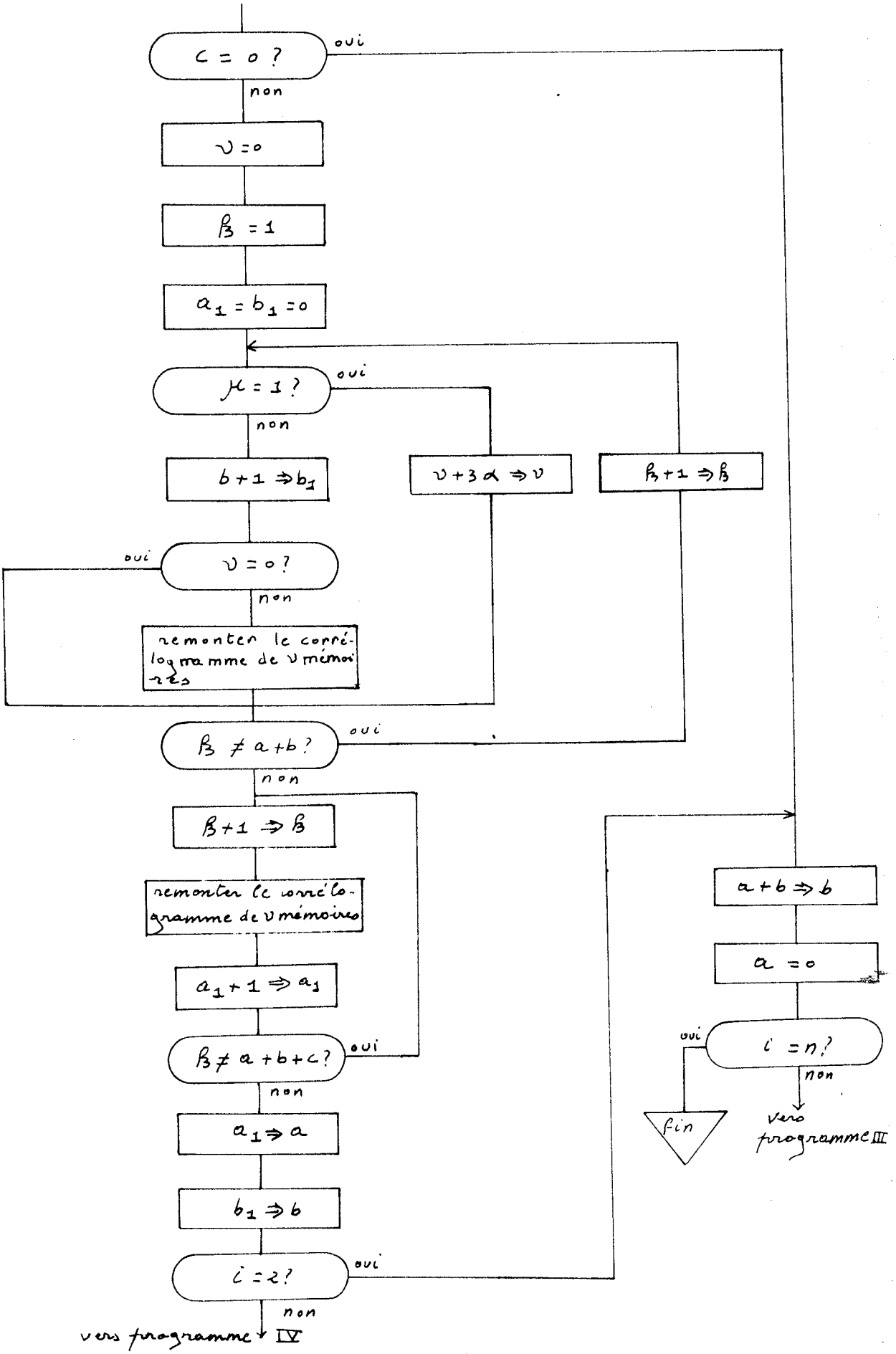
b_1 " " " B "

ν nombre de mémoires des corrélogrammes effacés.

i : numéro progressif du dernier mot étudié.

n : nombre de mots de la phrase.

Organigramme de la recirculation (programme 10)



D Algorithme permettant de déterminer la possibilité de liaison de deux corrélogrammes.

a Représentation de l'indice de saturation.

Cet indice occupe 3 bits que l'on désignera par a, b, c : a désignant le corrélateur, b le 2e corrélé, c le premier corrélé.

b Tableau de Boole- Il est évident que l'on peut avoir superposition seulement lorsque :

$f(a_1, b_1, c_1) = 111$ et $f(a_2, b_2, c_2) = 111$

(corrélogrammes saturés) et insertion lorsque :
$f(a_1, b_1, c_1) \neq 111$ et $f(a_2, b_2, c_2 \neq 111$

De plus il est évident que l'une au moins des variables a, b, c doit être égale à 1.

D'où le tableau de Boole réduit aux cas intéressants pour l'étude de l'insertion (F =1 s'il y a insertion possible, F =0 dans le cas contraire) dans lequel on note, en plus, l'indice de saturation du résultat éventuel dans le cas où F =1.

Il est à remarquer que l'on ne peut pas avoir F(a, b, c) = 011 puisque un réseau ayant un premier et un deuxième corrélé a obligatoirement un corrélateur (explicite ou non).

1er corrélé	2e corrélé	Résultat	F
a_1 b_1 c_1	a_2 b_2 c_2	a_3 b_3 c_3	
0 0 1	0 0 1		0
0 0 1	0 1 0	1 1 1	1
0 0 1	1 0 0	1 0 1	1
0 0 1	1 0 1		0
0 0 1	1 1 0	1 1 1	1
0 1 0	0 0 1	1 1 1	1
0 1 0	0 1 0		0
0 1 0	1 0 0	1 1 0	1
0 1 0	1 0 1	1 1 1	1
0 1 0	1 1 0		0
1 0 0	0 0 1	1 0 1	1
1 0 0	0 1 0	1 1 0	1
1 0 0	1 0 0		0
1 0 0	1 0 1	1 0 1	1
1 0 0	1 1 0	1 1 0	1
1 0 1	0 0 1		0
1 0 1	0 1 0	1 1 1	1
1 0 1	1 0 0	1 0 1	1
1 0 1	1 0 1		0
1 0 1	1 1 0	1 1 1	1
1 1 0	0 0 1	1 1 1	1
1 1 0	0 1 0		0
1 1 0	1 0 0	1 1 0	1
1 1 0	1 0 1	1 1 1	1
1 1 0	1 1 0		0

Posons $\beta_1 = f(a_1, b_1, c_1)$
$\beta_2 = f(a_2, b_2, c_2)$

Il y aura possibilité d'insertion lorsque $\beta_1 \cap \beta_2 = \emptyset$

Sinon, l'insertion sera impossible si

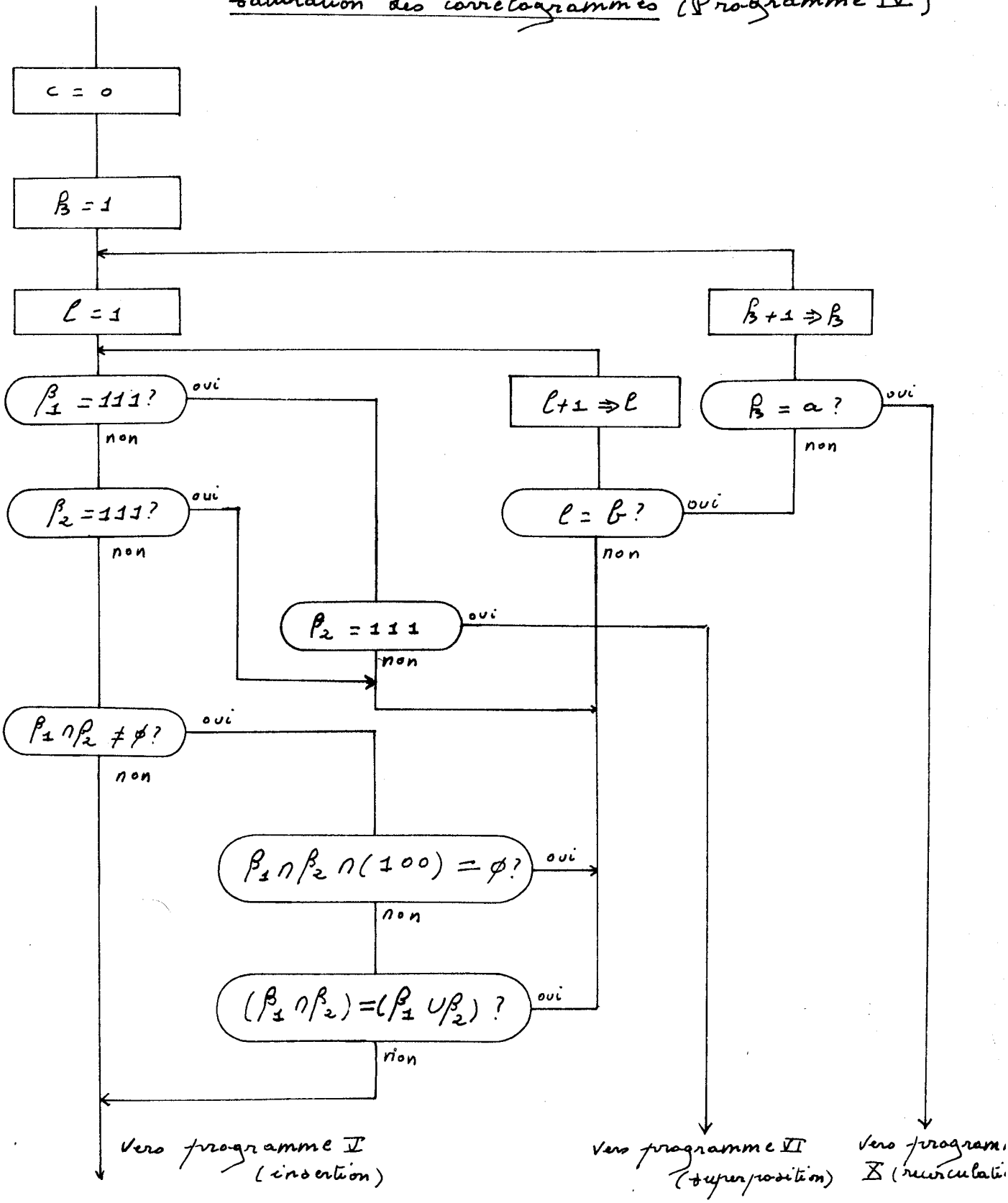
$\beta_1 \cap \beta_2 \cap (100) = \emptyset$ ou bien si

$(\beta_1 \cap \beta_2) = (\beta_1 \cup \beta_2)$.

Rappelons que, de plus l'insertion nécessite que β_1 et β_2 soient différents de 111, et que la superposition nécessite que $\bar{\beta}_1 = \bar{\beta}_2 = \emptyset$

c D'où l'organigramme ; les indices k et l repèrent les corrélogrammes respectivement dans les listes A et B.

Etude des possibilités de liaison d'après l'indice de saturation des corrélogrammes (Programme IV.)



B Codification des matrices-contrôle , mot et produit .

a Matrices-mot - Pour chaque ligne d'une matrice-mot, on donne des indications d'ordre morphologique syntaxique et sémantique qui ont, quel que soit le mot, la même structure. Ces informations seront fractionnées en 19 groupes correspondant chacun à une "colonne" de la matrice.

Ainsi une colonne correspondra à la catégorie lexicale, une autre au genre, une autre au nombre etc...

La structure de ces 19 types d'informations se trouvera dans toutes les matrices ; elle nécessite 5 mémoires pour être explicite, que nous désignerons par B C D E F.(Cf. page IV-26)

Si m est le nombre de lignes de matrice-mot, il faudra, pour donner toutes les indications relatives à ce mot, 6m + 3 mémoires ; les 2 premiers contiendront le mot lui même (en code I B M 7090) la 3e le numéro d'unité lexicale, la 4e contiendra m, l'indice corrélationnel et le numéro de fonction de corrélation correspondant à la première ligne de matrice, les 5 suivantes seront des mémoires du type B, C, D, E, F, la suivante contiendra l'indice de corrélation et le numéro de fonction de corrélation correspondant à la 2e ligne de la matrice etc...

b Matrice-contrôle - Chaque ligne de matrice-contrôle occupera 13 mémoires : 6 relatives au premier corrélé, 6 relatives au 2e corrélé et une relative à la corrélation. (Cf. page IV-28).

Les 6 mémoires relatives à un corrélé auront la structure suivante:

— Une mémoire A contient un code et peut contenir soit un numéro d'unité lexicale, soit un numéro de I_{cR} et sous-type ; le code permet d'exiger qu'un corrélé soit ou bien un mot (dont le numéro peut être précisé) soit un produit (dont le I_{cR} et sous-type peut être précisé).

— Les 5 mémoires du type B, C, D, E, F.

La mémoire relative à la corrélation donne les indications d'ordre, d'intervalle et enfin le I_{cR} et sous-type du résultat.

c ... Matrices-produit - Chacune d'elles occupent 7 mémoires : la prèmière contient le numéro du corrélogramme et, éventuellement les numéros des corrélogrammes premier et deuxième corrélé (cf. Codification des corrélogrammes) les 5 derniers sont du type B, C, D, E, F. (Cf. Page IV-27).

La deuxième mémoire, A, a une structure qui peut prendre 3 formes suivant que le corrélogramme correspondant a été obtenu par superposition (A_1), par insertion libre (A_3) ou par insertion saturée (A_2). Elle contient le nombre de mots du corrélogramme, le numéro du premier mot et celui du dernier, l'indice de corrélation et, éventuellement, le numéro du corrélateur.

Les règles permettant de déterminer les contenus des mémoires B, C, D, E, F d'une matrice-produit seront exprimées par une table contenant des codes désignant d'une part l'opération à effectuer, d'autre part le type d'information sur laquelle elle porte. (Matrices de reclassification).

7 -. Organisation des bandes d'entrée - Consultation du dictionnaire -

A — Trois bandes sont nécessaires pour stocker les informations nécessaires à l'analyse.

a ... Une bande contient les matrices — contrôle; les informations sont rangées sur cette bande, en un seul enregistrement ; sous la forme qui vient d'être vue.

b ... Une bande contient le texte à traduire, ce texte étant préédité manuellement (il n'y a aucun signe de ponctuation, le découpage en phrase est déjà fait). Un enregistrement contiendra sur les 6 premiers caractères le nombre de mots de la phrase, puis, à raison de 12 caractères par mot, les divers mots de la phrase ; chaque enregistrement sera donc composé de 2n + 1 mémoires, n étant le nombre de mots de la phrase à analyser.

c ... Un dernière bande contient le dictionnaire de forme ; chaque enregistrement correspond à un mot ; la codification correspond à ce qui vient d'être dit sur les matrices-mot.

La consultation du dictionnaire est réduite au minimum : aucun tri n'est prévu car le dictionnaire ne contient que 104 formes, elle se fait donc mot par mot.

8-. Posiibilités et limites du programme

Les données linguistiques dont on dispose actuellement sont extrèmement réduites : le dictionnaire est très rudimentaire (il contient 104 formes de la langue anglaise choisies parmi les plus courantes) et le nombre de corrélateurs explicites est très faible (7 sur la centaine actuellement recensée). Il ne faut donc pas oublier que ce programme n'est pas destiné à être exploité pour de véritables traductions mécaniques, mais qu'il doit servir de guide aux études ultèrieures.

Le programme lui-même a été conçu de la façon la plus ouverte possible : il pourrait traiter des phrases ayant un nombre élevé de mots (127), utiliser 63 corrélateurs explicites ; pour une utilisation de ce programme avec un dictionnaire plus vaste, il suffirait de refaire la consultaion de dictionnaire qui, actuellement est extrèmement lente.

Le programme doit être d'ici peu continué afin de traduire en latin le texte d'entrée ; il faut pour cela un programme de transformation des réseaux corrélationnels puis un programme de flexion de la langue cible ; la formalisation de ces transformations est en cours d'étude.

L'arborescence des réseaux est actuellement extrêmement grande, et pour une même phrase, on trouve plusieurs corrélogrammes ayant des structures équivalentes et conduisant à une même traduction en langue cible il faudrait donc restreindre certaines règles, en particulier ne pas faire de superposition systématique, et ne pas reclassifier chaque corrélogramme saturé par 4 corrélogrammes (il semble qu'il serait suffisant de créer deux corrélogrammes lors de cette reclassification, mais il n'est pas encore possible de l'affirmer avec certitude).

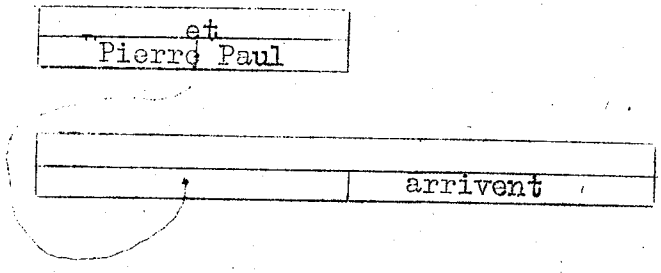
9-. Remarques sur les corrélogrammes - Relation avec les graphes de Tesnières et de Chomsky.

L'étude structurale d'une phrase à l'aide de corrélogrammes a été guidée par celle de la compréhension par l'homme du sens d'une phrase. Le point de départ de la théorie de la corrélation du professeur Ceccato est d'ordre psychologique, et la traduction mécanique n'est qu'une application de cette théorie.

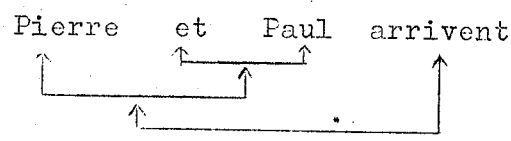
Cependant, il est possible, connaissant le corrélogramme correspondant à une phrase, de retrouver la structure de Chomsky, et de là, celle de Tesnières (En effet Lecerf a déjà démontré l'équivalence de ces deux types de graphes).

On retrouve en effet dans le corrélogramme les liaisons entre deux mots ou groupes de mots avec leurs niveaux : il suffit de faire la convention suivante : lorsqu'un rectangle contient un corrélateur explicite, il donne lieu à deux liaisons, celle de plus bas niveau étant faite avec le 2e corrélé.

La taille respective de deux rectangles liés donne le syntagme prépondérant.

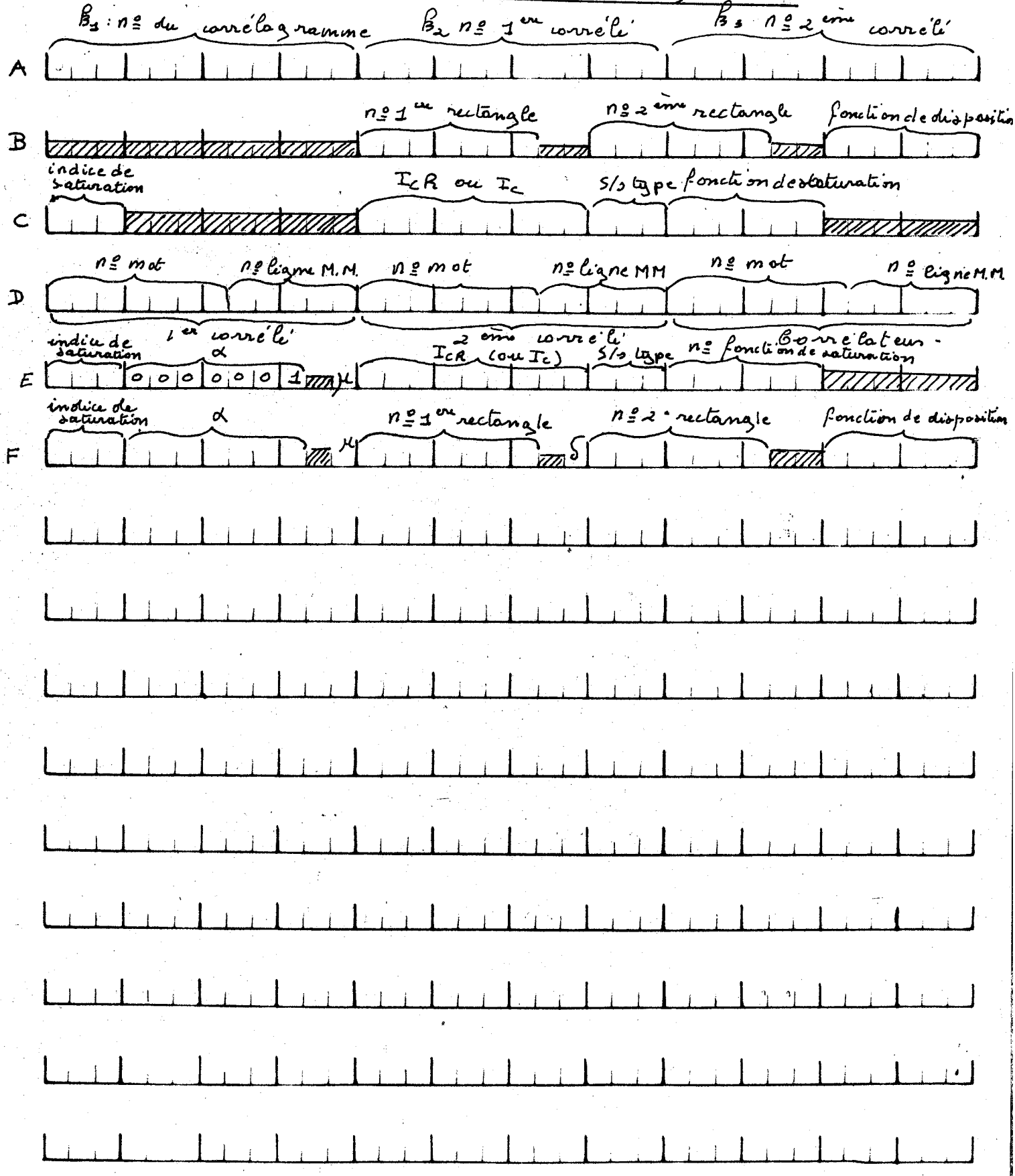


Corrélogramme de Ceccato.

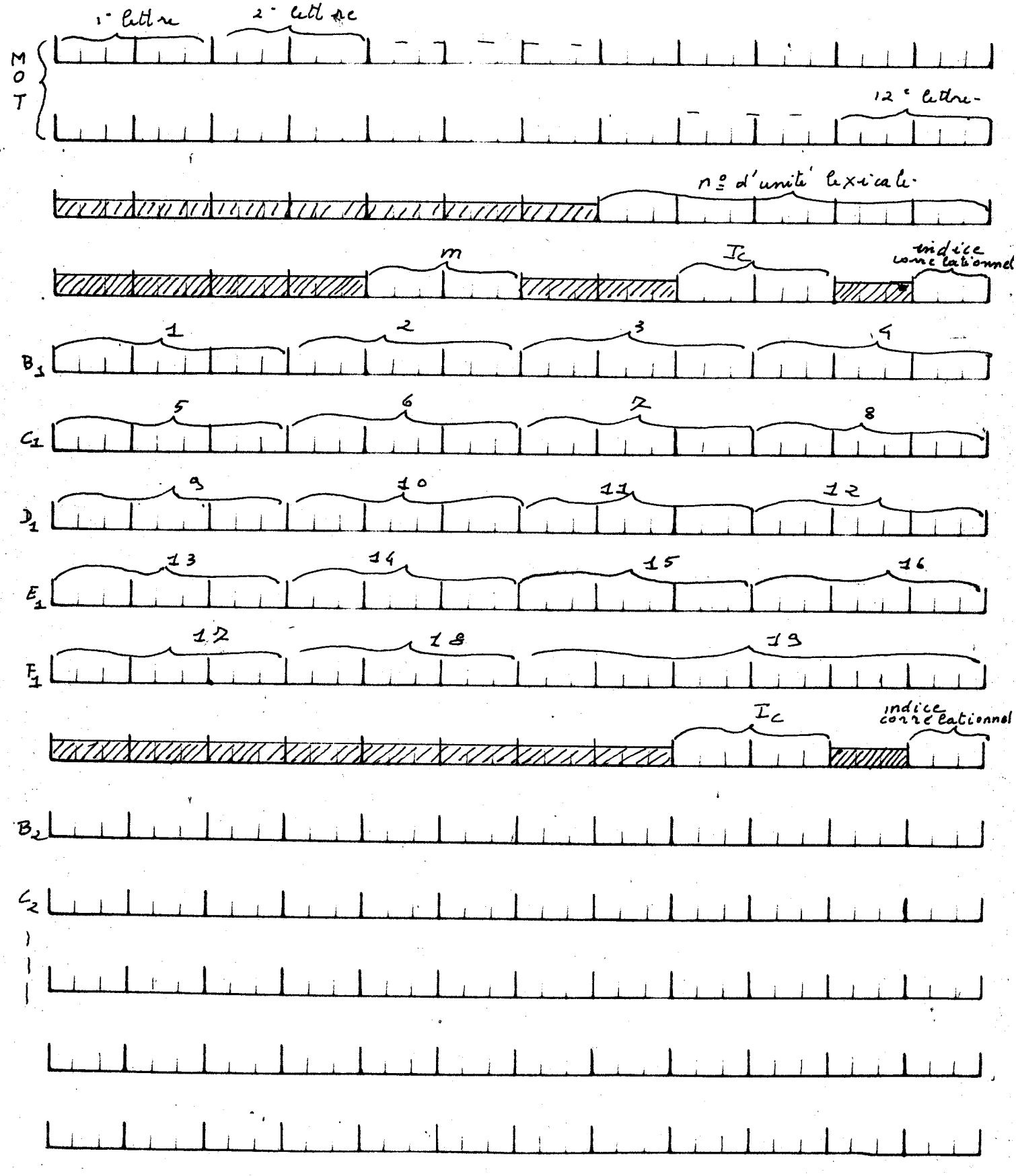


Graphe de Chomsky.

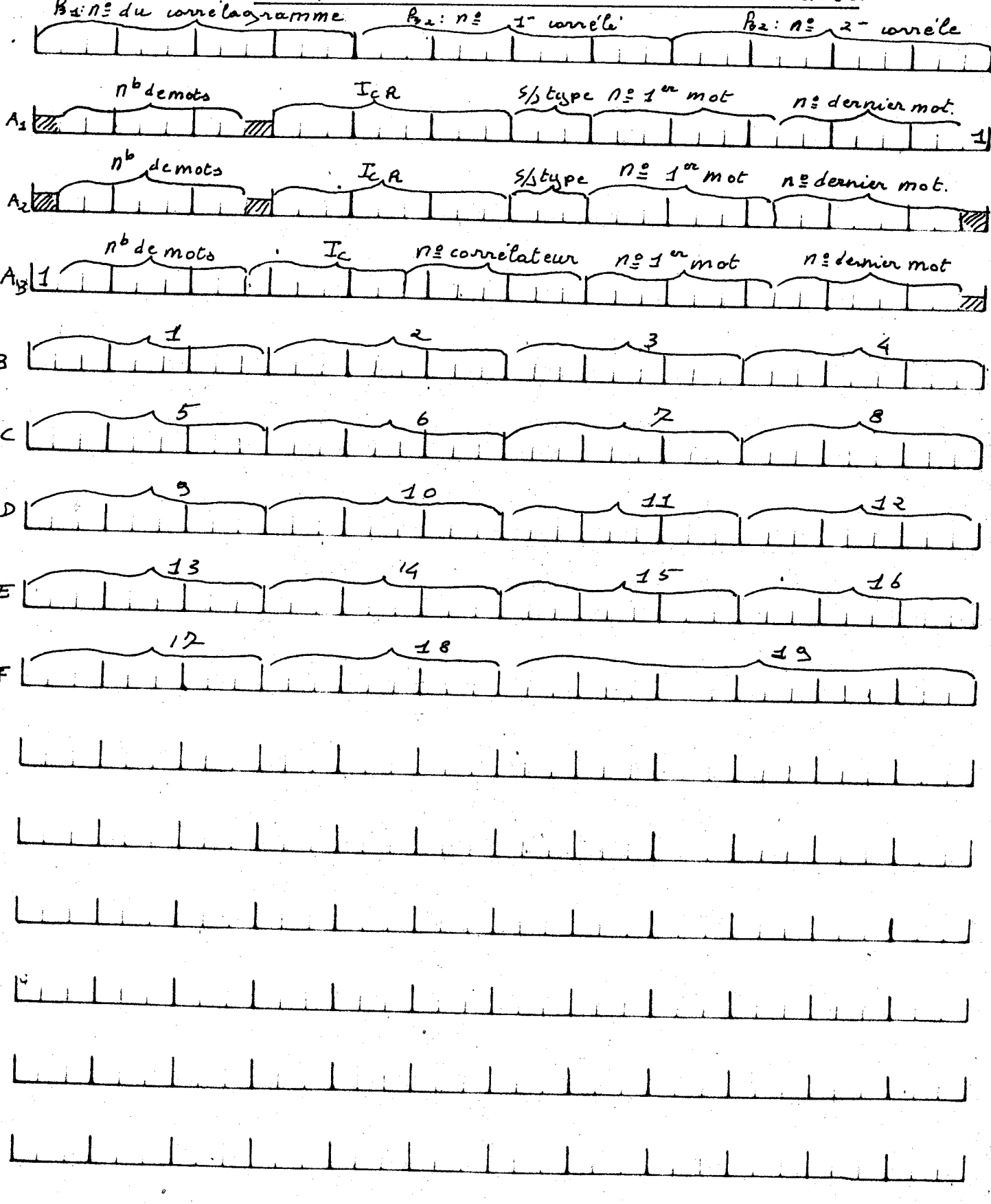
codification des corrélogrammes



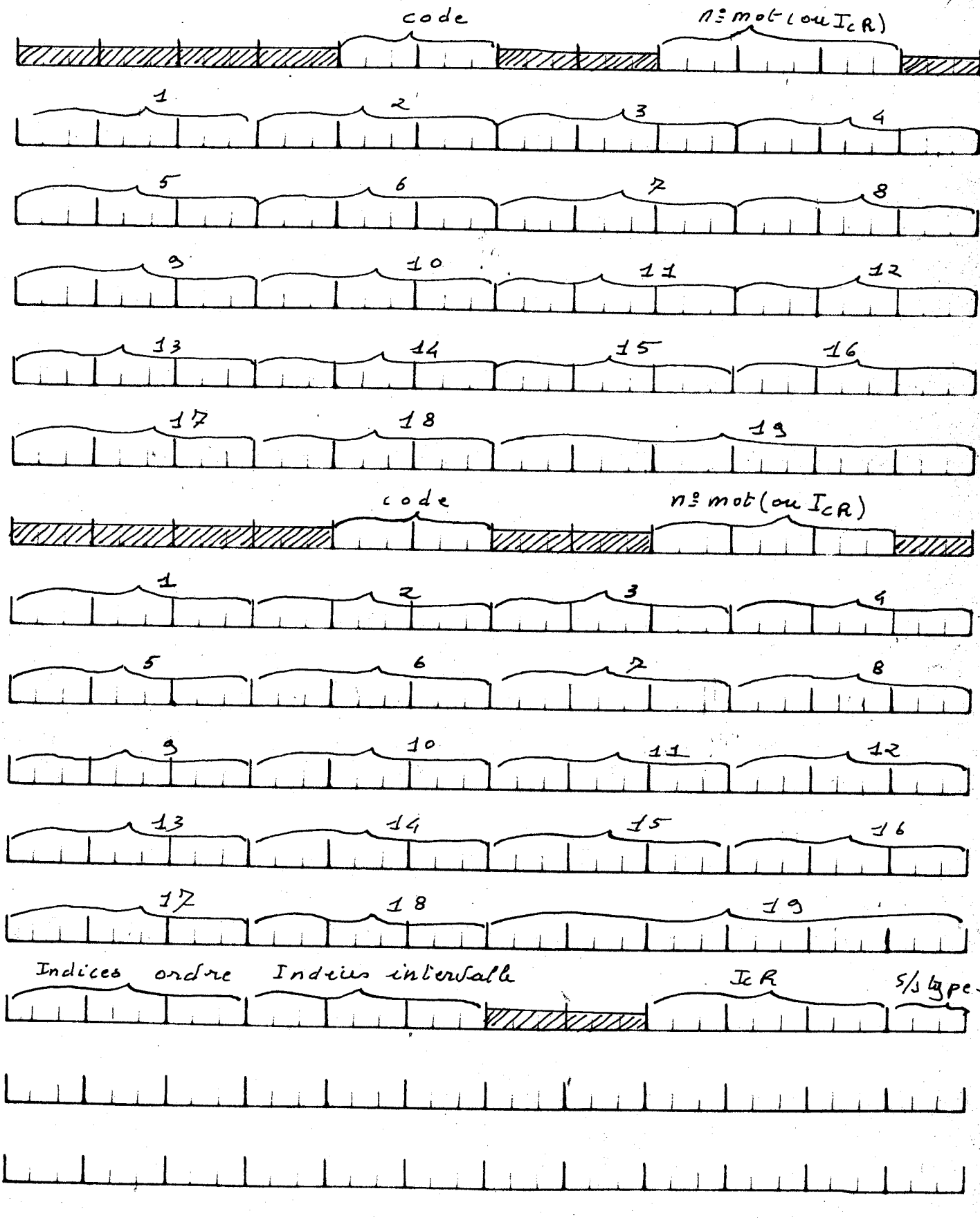
codification des matrices - Mot



codification des matrices Produit.



codification d'une ligne de matrice controle.



C H A P I T R E V

ASPECT PRATIQUE DES ETUDES SYNTAXIQUES AUTOMATIQUES. STRATEGIES ENVISAGEES.

Les données de bases de l'algorithme syntaxique sont, comme nous l'avons vu des familles d'homographes internes.

Soit alors une phrase de L_1 que l'on désire analyser; le programme d'analyse morphologique va faire correspondre à chaque occurrence une ou plusieurs familles d'homographes internes.

1-. Résolution des homographies externes.

Le premier problème qui se pose sera comme il a été dit au chapitre I la résolution des cas d'homographie externe : dans ce cas à une occurrence de L_1 correspond plusieurs familles d'homographes internes ayant des ku différents ; cette résolution est nécessaire pour pouvoir appliquer l'algorithme décrit au chapitre III.

Soient deux occurrences consécutives auxquelles correspondent respectivement les catégories lexicales $Ku_1, Ku_2 .. Ku_n$ et $Kv_1, Kv_2 .. Kv_n$; une étude linguistique de L_1 nous permet d'affirmer que la juxtaposition de Ku_i et Kv_j est impossible ; d'où un algorithme permettant la levée des homographies externes.

Construisons la table de tous les couples Ku Kv possible : une exploration de cette table dans tous les cas d'homographie externe doit permettre de résoudre ce type d'ambiguité.

Prenons un exemple ; soit la phrase :

la	porte	était	ouverte
(1)	(2)	(3)	(4)

Il y a homographie externe pour "la" qui peut être article défini ou pronom personnel, et pour "porte" qui peut être verbe personnel ou substantif. A priori il y a donc 4 cas possibles :

	1	2	3	4
1 -	article	verbe	verbe	adjectif
2 -	"	substantif	"	"
3 -	pronom	verbe	"	"
4 -	"	substantif	"	"

Par consultation de la table envisagée ci-dessus, on éliminera successivement les cas 1, 4 et 3.

Un tel procédé doit permettre de lever la majorité des cas d'homographie externe.

S'il reste encore des ambiguités, il faudra alors se ramener au cas considéré au chapitre II, c'est à dire le cas où il n'y a plus d'homographie externes ce qui nécessitera dans la suite plusieurs études syntaxiques.

Il est à remarquer que la table des couples Ku, Kv possibles est une partie (correspondant à I = o) de celle dont on a parlé au chapitre II lors de la recherche des fonctions ψ (Ku, Kv) = Kw, il serait donc possible de noter, lors de l'étude des homographies externes, les ψ_q correspondant à deux Ku et Kv possibles.

2-. Etude des cas continus et discontinus, projectivité

Considérons le schéma en arbre correspondant à une phrase analysée :

S'il n'y a de liaisons qu'entre des syntagmes consécutifs, on dira que l'on a une arborescence continue et que l'on a une arborescence

discontinue dans le cas contraire;on dira encore que, dans le premier cas, on a affaire à une phrase projective et à une phrase non projective dans le second.

Ce dernier cas est, tout au moins du point de vue pratique, assez génant ; en effet la fonction ψ (Ku,Kv) = Kw qui dépend, dans le cas continu, des seules catégories lexicales des familles d'homographes internes considérées, Ku et Kv, dépend en plus dans le cas discontinu de l'amplitude de l'intervalle et des catégories lexicales des familles d'homographes qui le constituent.

Il semble qu'il n'y ait jamais d'intervalles supérieurs à 1, c'est à dire que, dans le cas général, les fonctions ψ dépendront de 4 paramétres : Ku, Kv, I pouvant prendre les valeurs 0 ou 1, et Kx catégorie lexicale de la famille d'homographes internes se trouvant dans l'intervalle.

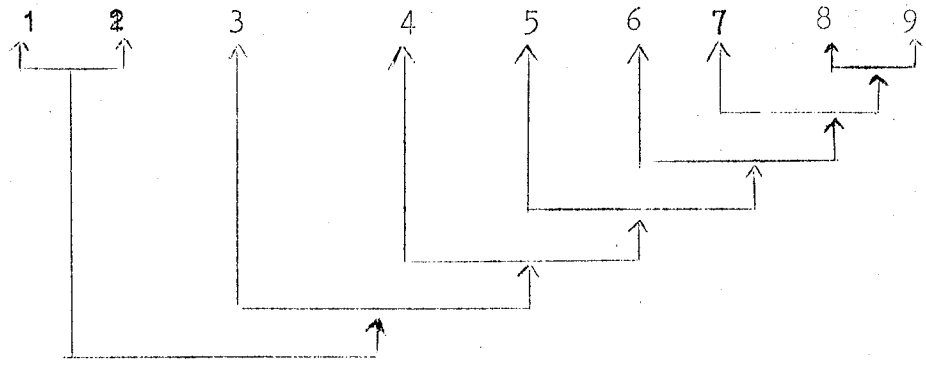
On peut envisager plusieurs méthodes pour l'étude des cas discontinus :

- la première (Mm HIRSCHBERG) consiste à rechercher en priorité les liaisons correspondant au cas continu, puis à étudier les liaisons discontinues lorsque l'on n'a pas trouvé de syntagme décrivant la phrase entière ; ce procédé est séduisant, mais dans certains cas (évidemment en nombre restreint) d'ambiguité sémantique il conduit à l'élimination de solutions valables;

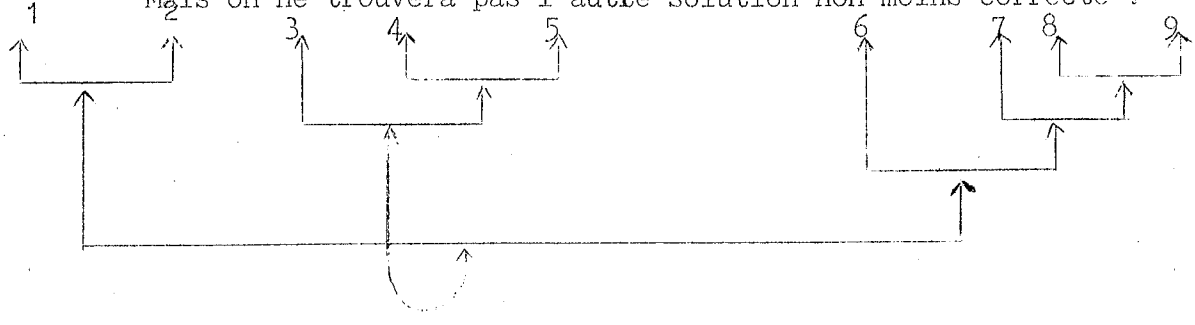
Prenons l'exemple suivant : (Mme HIRSCHBERG : discussion sur l'hypothèse de projectivité, exemple 25)

le	prisonnier	sort	du	tribunal	gardé	par	deux	gendarmes
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)

en appliquant cette méthode on trouvera la solution suivante, qui est correcte :



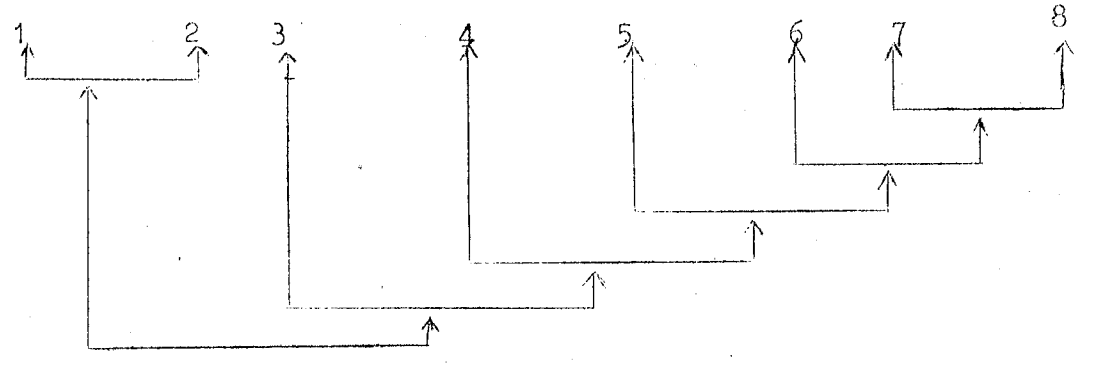
Mais on ne trouvera pas l'autre solution non moins correcte :



Cette solution est applicable seulement lorsque la phrase à étudier est non ambigue au moins sémantiquement, mais semble difficile à mettre en oeuvre avec le programme d'analyse syntaxique et sémantique envisagé; en effet si l'on a la phrase suivante :

le	prisonnier	sort	du	tribunal	habillé	en	rouge
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

Le résultat syntaxique serait :



résultat évidemment faux mais qui serait admis si l'on fait appel au programme d 'analyse sémantique seulement pour lever les ambiguïtés syntaxiques (cf. role du programme d'analyse sémantique page I-3)

Une deuxième méthode, pratiquement plus longue à mettre en oeuvre, consistera à chercher systématiquement les cas discontinus en explorant une table donnant toutes les possibilités de fonctions ψ pour I-1 les variables intervenant étant alors Ku, Kv et Kx catégorie lexicale de la famille d'homographes internes se trouvant dans l'intervalle. Cette méthode permet d'étudier tous les cas de discontinuité y compris ceux présentant une ambiguïté sémantique. Il est d'autre part possible de ne pas essayer tous les triplets Ku, Kx, Kv car les cas discontinus sont assez rares, et ne seront possibles qu'avec un nombre restreint de couples Ku, Kx ou Kx, Kw.

b-. Elimination de certaines discontinuités apparentes.

Certaines discontinuités ne sont qu'apparentes : c'est le cas, en français de bien des négations telles que ne....pas ; ne.....que ne....point etc... il sera très simple de se ramener, dans ce cas particulier, à une structure continue : il suffit pour cela de ne tenir compte que de l'un des marquants. Ainsi dans le cas de l'une des négations françaises vues ci-dessus, On ne considerera que le marquant "ne" qui est indispensable à la négation (alors que le 2° marquant, pas, point, ne l'est pas)

c-. Etude des séparateurs.

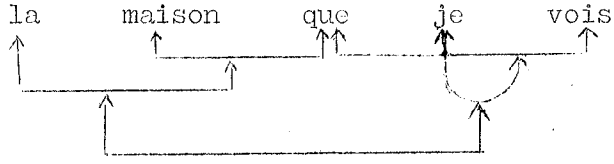
Nous allons chercher à morceler l'étude syntaxique d'une phrase en la ramenant à l'étude d'éléments de moindre dimension ; chacun de ces éléments ayant une unité syntaxique ; pour cela il faudra rechercher des séparateurs, c'est à dire des syntagmes élementaires marquant le

début ou la fin d'un élément d'étude.

Parmi les séparateurs, nous distinguerons les séparateurs inconditionnels (certains signes de ponctuation, les pronoms relatifs) et les séparateurs conditionnels qui ne seront séparateurs que dans certains cas. La virgule est, par exemple un séparateur conditionnel : en effet on ne la considérera pas comme séparateur lorsqu'elle marque une énumération.

Suivant qu'un séparateur marque le début ou la fin d'un élément d'étude on l'appellera "séparateur à gauche" ou "séparateur à droite" certains mots (tel le pronom relatif "que") sont à la fois séparateur à droite et séparateur à gauche. Il faudra alors attribuer deux liaisons à un tel mot.

Exemple :



Chaque séparateur sera caractérisé par une portée à droite ; cette notion est très importante pour la détermination des élements d'étude souvent la portée d'un séparateur s'étendra jusqu'au séparateur le plus proche.

Ainsi, dans la phrase "la maison que je vois est grande" la portée à droite de "que" considéré comme séparateur à gauche s'étend jusqu'à "vois" compris.

5-. Recherche de groupements projectifs

A l'aide des séparateurs nous avons déterminé pour une phrase, un certain nombre d'éléments correspondant en quelque sorte à une proposition ; nous allons maintenant, par une étude plus fine, chercher à décomposer cet élément en groupes projectifs ou prédicats ; un élément pouvant être décomposé, dans le cas le plus complexe, en prédicat nominal sujet, prédicat verbal et un un certain nombre de prédicats complementaires ; ainsi l'élément "que je vois" comprend 3 prédicats :

que	prédicat complementaire
je	prédicat nominal sujet
vois	prédicat verbal

Une fois déterminés ces prédicats, il faudra pour chacun d'eux faire l'étude de structure syntaxique, ensuite lier les divers prédicats appartenant à une même proposition, et enfin lier les propositions entre elles.

L'étude de la structure d'un prédicat sera très simple: elle se ramène à la recherche des concaténations valables par cohérence syntaxique. Le problème le plus difficile à résoudre, et celui qui sera à l'origine des ambiguïtés syntaxiques, sera la recherche des prédicats.

Ainsi la phrase "le prisonnier sort du tribunal habillé en rouge" forme un seul élément. Elle est décomposable en 3 ou 4 prédicats :

1- Prédicat sujet : "le prisonnier"
" verbal: "sort"
" complémentaire : "du tribunal"
" " : "habillé en rouge"

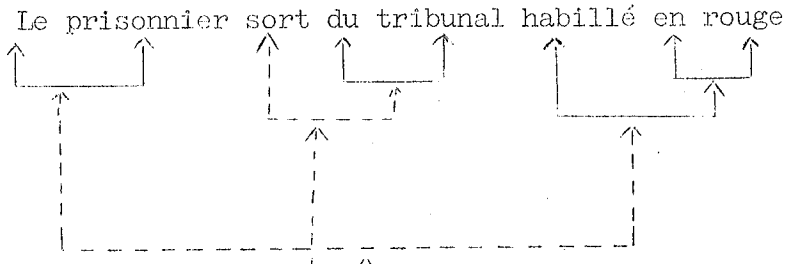
2- Prédicat sujet "le prisonnier"

" verbal "sort"

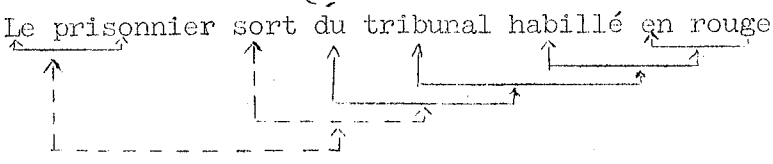
" complémentaire "du tribunal habillé en rouge"

Ce qui conduit à deux structures syntaxiquement valables :

1.



2.



5- Liaisons multiples

Nous avons vu qu'il était nécessaire de mettre plusieurs liaisons sur certains séparateurs tels que les pronoms relatifs ; il serait intéressant de généraliser ce procédé à certaines autres classes telles que les adjectifs posséssifs, les pronoms personnels. En effet il serait bon de lier un adjectif possessif au possesseur et au possédé ; mais cela sort du domaine de la syntaxe pour entrer dans celui de la sémantique.

6- Sens d'étude d'une phrase

Deux modes d'études peuvent être envisagés : soit de gauche à droite (CECCATO) soit de droite à gauche (YNGVE). Le premier procédé, qui semble le plus naturel parce que correspondant à l'ordre de la parole et, semble-t-il, de la pensée n'est peut être pas celui qui conduit à l'analyse syntaxique la plus rapide. En effet il est remarquable que dans la majorité des cas une fonction φ_p lie une famille d'homographes internes (élementaire ou non) à la famille d'homographes internes située immédiatement à sa gauche. Il semble donc opportun d'adopter le deuxième procédé, au moins pour l'étude des prédicats.

8- Retour sur la recherche de la structure d'un prédicat.

Il sera possible, le découpage en prédicats supposé fait, de mener globalement l'étude de ces derniers. Pour cela (cf paragraphe 5) on dressera une table donnant toutes les concaténations de Ku valables.

Cette table est évidemment assez vaste, mais il est possible d'une part de la raccourcir notablement, d'autre part de réduire au minimum son temps d'exploration par un rangement en arbre.

Ce procédé doit permettre une étude rapide et simple des prédicats, qui se fera en 2 étapes :

-cheminement le long de l'arbre afin de voir si la concaténation des Ku est valable

-étude de la cohérence syntaxique.

9- Rapport entre l'analyse syntaxique et l'analyse sémantique.

Alors qu'il est possible de faire l'étude morphologique d'un texte entier avant de commencer l'analyse syntaxique, il n'en sera pas de même pour les études syntaxiques et sémantiques. En effet cette dernière à deux roles principaux :

- faire un choix dans le cas de polysémies
- éliminer certaines structures syntaxiquement exactes et remplacer les fonctions syntaxiques par des fonctions sémantiques.

Ce dernier rôle ne peut être joué qu'en liaison étroite avec l'analyse syntaxique

On peut envisager la liaison entre les deux analyses de deux façons distinctes :

- soit faire l'analyse syntaxique complète d'une phrase, puis, si plusieurs structures sont trouvées, faire l'analyse sémantique

destinée à éliminer certaines d'entre elles.

- soit faire appel au programme d'analyse sémantique, dès que l'on trouve une ambiguïté syntaxique.

Ce dernier procédé, à condition de ne pas être appliqué à un niveau trop bas, semble préférable . Il permet de ne pas traiter certaines structures sémantiquement fausses, ce qui réduit les données et doit raccourcir le temps d'éxécution de l'analyse syntaxique.

Pour mettre en oeuvre le premier procédé, il faudrait noter, chaque fois que l'on trouve plusieurs structures possibles, l'endroit où se trouve cette ambiguité, et son niveau.

L'analyse sémantique remplacera les fonctions syntaxiques φ_p par des fonctions sémantiques ; à une fonction φ_p correspond au moins une fonction sémantique ; afin de raccourcir cette analyse, le programme d'analyse syntaxique indiquera dans les φ_p si à une fonction syntaxique correspond une ou plusieurs fonctions sémantiques.

B I B L I O G R A P H I E

S. CECCATO — La traduzionne Meccanica (Automazione e Automatismi n°2. mars avril 1958)
Operational linguistics and translation. (Linguistic Analysi and Programming for Mechanical Translation).

N. CHOMSKY — Syntactic Structures

D.G. HAYS — Basic principles and technical Variations in sentence structures détermination (4ème London symposium on information théory, 2 Sept. 1960)

L. HIRSCHBERG — Discussion sur l'hypothèse de projectivité.

S. KUNO — A préliminary approach to Japanese-English Automatic Translation (first International Congress on Machine translation of Language - Teddington, September 1961)

Y. LECERF — Le programme des conflits 'Enseignement préparatoire aux techniques de la documentation automatique (rapport Euratom).

E. MAJETTI — [illegible] représentation of operational procedure (linguistic Analysis and programming for mechanical Translation).

A.G. OETTINGER — Progress report NSF4 (computation laboratory Harvard university).

I. RHODES — A new approach to the mechanical syntactic analysis of Russian (Mechanical Translation, n°6, November 1961)

L. TESNIERES — Elements de syntaxe structurale.

B. VAUQUOIS — Communication au 1.er congrès de l'AFCAL, Grenoble Octobre 1960)
Document interne du CETAG G.100 C.

G. VEILLON — Consultation d'un dictionnaire et analyse morphologique en traduction automatique (thèse de 3ème cycle Université de Grenoble, Juin 1962).

J. VEYRUNES — Document interne du CETAG G100 C

V. YNGVE — A model and an hypothesis for language structure (Proceding of the American Philosophical Society, Mars 1960)

B. ZONTA — An example of procedure (linguistic Analysis and programming for mechanical Translation).

T A B L E des M A T I E R E S

VU,

Grenoble, le

Le Président de la Thèse

VU,

Grenoble, le

Le Doyen de la Faculté des Sciences

VU et permis d'imprimer,

Le Recteur de l'Académie de Grenob