

## PLAN

- I- BUT ET LIMITES DE CETTE NOTATION
- II- STRUCTURE DU LANGAGE PIVOT
  - a) Elément du langage
  - b) Lexis et énoncés élémentaires
  - c) Composition des lexis et des énoncés élémentaires
- III- ACCES AUTOMATIQUE AU LANGAGE PIVOT
  - a) Structure syntaxique initiale
  - b) Description de la grammaire d'étiquetage (M3)
- IV- APPLICATIONS
  - a) Utilisation dans un programme de traduction automatique
  - b) Nature du lexique
  - c) Notation de la structure syntaxique
  - d) La grammaire de production
  - Autres applications

Annexes

-I- BUT ET LIMITES DE CETTE NOTATION

On appelle "expression" la manière dont est rédigé un texte (ou discours) dans une langue naturelle. Les textes ainsi articulés en phrases, les phrases en propositions, etc... jusqu'au niveau morphémique obéissent aux règles de la morphologie et de la syntaxe de la langue dans laquelle ils sont exprimés. En général, les choix multiples que l'on peut faire sur les constructions morphologiques et syntaxiques d'une part et sur le lexique d'autre part, conduisent à de nombreuses expressions différentes d'une même signification. L'optique des modèles texte-sens et sens-texte consiste à déterminer une notation au moyen de laquelle la signification des textes est enregistrée indépendamment des particularités grammaticales et lexicales de la langue d'expression. Une telle notation serait la notation sémantique idéale. On peut admettre qu'une notation effective est d'autant plus approchée de cette notation idéale qu'elle permet de reconnaître (et d'engendrer des familles plus vastes d'expressions ayant des significations jugées équivalentes.

Ce problème a été abordé dans le cadre d'un modèle sens-texte par A. JOLKOVSKY et I. MELTCHOUK (1), (2). Il trouve aussi son importance dans le modèle réversible de S. LAMB (3). Enfin il a été l'objectif principal des travaux du C.E.T.A. dans un but de traduction automatique (4), (5) et continue de se poser pour d'autres applications.

La réalisation d'une notation effective est soumise à un certain nombre de contraintes qui elles-mêmes imposent des limites.

Un premier type de contraintes est lié à l'aspect "génératif" d'un modèle sens-texte. La notation est alors liée aux possibilités de dérivation des structures syntaxiques et lexicales dans une langue donnée.

Un deuxième type de contraintes, plus sévère que le précédent, est lié à l'aptitude à reconnaître par un procédé automatique l'équivalence proposée entre deux expressions et à calculer leur formulation. En effet, l'ambiguïté de signification des expressions est plus difficile à résoudre que la multiplicité des expressions équivalentes à construire.

Enfin, l'application d'une telle notation en vue d'une activité particulière fait apparaître des contraintes très différentes. Les conditions d'utilisation en traduction automatique, en paraphrasage dans une langue, en communication homme-machine en langue naturelle, etc... obéissent à des critères différents. Ainsi, la solution de certaines ambiguïtés se révèle inutile en traduction automatique lorsque ces ambiguïtés se retrouvent dans les expressions des langues en jeu ; dans un problème de communication homme-machine de telles ambiguïtés seraient inadmissibles, en contre partie le système de communication peut prévoir une interrogation de l'homme par la machine pour déterminer la solution alors que ce "feed back" ne peut être utilisé en traduction automatique.

La notation exposée ici, appelée "langage pivot I" a été réalisée dans un but de traduction automatique. Il est donc nécessaire d'accéder automatiquement à cette notation à partir du texte. Cet aspect est exposé en détail dans la partie III.

Pour cette première expérience les contraintes supplémentaires suivantes ont été acceptées :

- a) l'Unité d'expression est déterminée par la "phrase syntaxique"
- b) l'équivalence des expressions ayant même signification est davantage limitée par les possibilités de transformations lexicales que par la variation des constructions syntaxiques.

## -II- STRUCTURE DU LANGAGE PIVOT

### a) Élément du langage

Il semble qu'une notation cohérente et complète puisse être obtenue à l'aide des trois types d'éléments, unités lexicales variables, relations.

### - Lexique

Une notation indépendante des modes d'expressions des langues naturelles doit avoir son propre lexique. Dans la première étape, présentée ici, le langage pivot ne possède pas sa propre lexicographie. Ainsi, dans l'application à la traduction automatique, le passage lexicographique de la langue source à la langue cible est la seule opération de "transfert" c'est-à-dire une opération qui n'est ni d'analyse ni de synthèse.

Chaque élément lexicographique du langage pivot est donc donné par le couple "unité lexicale russe - ensemble de mots français dérivant d'une même signification". (voir à titre d'exemple les tableaux en partie III).

Enfin, l'application d'une telle notation en vue d'une activité particulière fait apparaître des contraintes très différentes. Les conditions d'utilisation en traduction automatique, en paraphrasage dans une langue, en communication homme-machine en langue naturelle, etc... obéissent à des critères différents. Ainsi, la solution de certaines ambiguïtés se révèle inutile en traduction automatique lorsque ces ambiguïtés se retrouvent dans les expressions des langues en jeu ; dans un problème de communication homme-machine de telles ambiguïtés seraient inadmissibles, en contre partie le système de communication peut prévoir une interrogation de l'homme par la machine pour déterminer la solution alors que ce "feed back" ne peut être utilisé en traduction automatique.

La notation exposée ici, appelée "langage pivot I" a été réalisée dans un but de traduction automatique. Il est donc nécessaire d'accéder automatiquement à cette notation à partir du texte. Cet aspect est exposé en détail dans la partie III.

Pour cette première expérience les contraintes supplémentaires suivantes ont été acceptées :

- a) l'Unité d'expression est déterminée par la "phrase syntaxique"
- b) l'équivalence des expressions ayant même signification est davantage limitée par les possibilités de transformations lexicales que par la variation des constructions syntaxiques.

## -II- STRUCTURE DU LANGAGE PIVOT

### a) Élément du langage

Il semble qu'une notation cohérente et complète puisse être obtenue à l'aide des trois types d'éléments, unités lexicales variables, relations.

### - Lexique

Une notation indépendante des modes d'expressions des langues naturelles doit avoir son propre lexique. Dans la première étape, présentée ici, le langage pivot ne possède pas sa propre lexicographie. Ainsi, dans l'application à la traduction automatique, le passage lexicographique de la langue source à la langue cible est la seule opération de "transfert" c'est-à-dire une opération qui n'est ni d'analyse ni de synthèse.

Chaque élément lexicographique du langage pivot est donc donné par le couple "unité lexicale russe - ensemble de mots français dérivant d'une même signification". (voir à titre d'exemple les tableaux en partie III).

Quelle que soit la notation lexicale (propre ou empruntée) les unités sont distribués suivant l'une ou l'autre des classes suivantes :

- éléments à valeur prédicative  
(verbes, substantifs verbaux, adjectifs, prépositions, conjonctions, etc...)
- éléments à valeur non prédicative  
(en général les mots descripteurs)

#### - Variables

On range dans ce type d'éléments les informations qui servent à l'actualisation du discours (passage d'une lexis à un énoncé élémentaire, passage de l'articulation des lexis à l'actualisation de cette articulation, etc...)

Dans l'application à la traduction automatique, ces informations sont appelées "variables persistantes" car elles sont déduites de l'expression du texte en langue source et doivent être exprimées dans la langue cible pour conserver le sens.

Par exemple la variable "énonciation" dont les valeurs sont "affirmative" et "négative" ; de même le temps (temps réel et non temps syntaxique), aspect, etc...

#### - Relations

Ce sont des métaprédicats du langage pivot dont certains établissent la place des arguments des prédicats et les autres indiquent les relations entre lexis ou leurs arguments.

Toutes les relations utilisées sont des métaprédicats à 2 places d'arguments.

#### b) Lexis et énoncés élémentaires

On admet que, dans le discours, la construction la plus simple est l'énoncé élémentaire représenté par un prédicat (extrait du lexique) muni de ses arguments (extraits du lexique) et des variables portant sur le prédicat et les arguments.

En réalité, on s'intéresse d'abord à une construction plus squelettique qui est la "lexis". la définition de la lexis a déjà été donnée par (6),(7). Rappelons cependant qu'une unité de lexique

à valeur prédicative représente une "notion". Par exemple l'unité "LIRE" représente la notion de lire qui par la suite peut être actualisée en prédicat "lire" ou "ne pas lire", à divers temps, aspect, etc... "LIRE" possède deux places d'argument, soit LIRE (x,y) où x et y sont des variables formelles qui parcourent l'ensemble du lexique. Si l'on prend pour x l'unité lexicale "secrétaire" et pour y l'unité lexicale "journal" on obtient la lexis :

LIRE (secrétaire, journal)

qui peut donner naissance aux énoncés élémentaires exprimés en français par

"Le secrétaire n'a pas lu les journaux"

ou encore

"La secrétaire est en train de lire ce journal"

ou encore

"la lecture des journaux par les secrétaires" etc...

suivant les valeurs des variables d'actualisation.

Pour construire la lexis dans le langage pivot on dispose des relations qui placent les arguments dans une notion :

Soit ACT n (a,P) où n = 1, 2 ou 3

a est une unité du lexique

P est une unité du lexique à valeur prédicative (notion)

Ainsi ACT1 (a,P(x,y)) = P(a,y)

ou ACT3 (c,P(x,y,z)) = P(x,y,c)

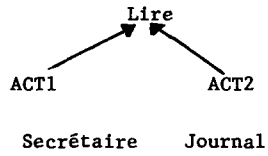
on obtient la lexis LIRE (secrétaire, journal) au moyen des deux relations

et ACT1 (secrétaire, Lire(x,y))

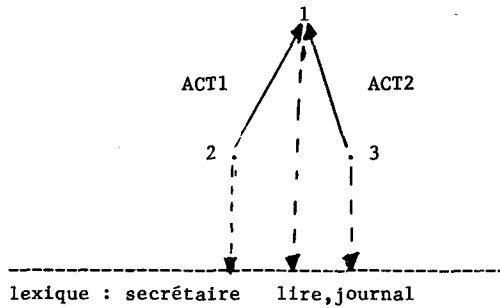
ACT2 (journal, Lire(x,y))

Il est commode de représenter ces relations par un graphe où chaque arc caractérise une relation, l'origine étant le premier argument et l'extrémité étant le deuxième argument de la relation.

On obtient ainsi :

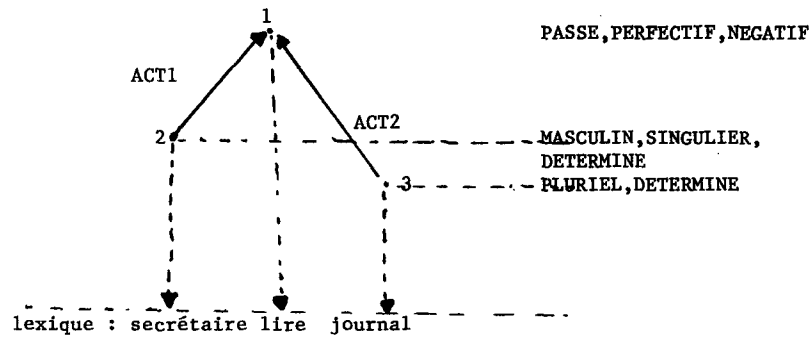


Afin de mettre en évidence d'une part le graphe des relations et d'autre part le lexique on définit une application  $\Delta$  des sommets du graphe dans le lexique ; on obtient alors la représentation de la lexis par :



Enfin, l'énoncé élémentaire

"Le secrétaire n'a pas lu les journaux" est noté en ajoutant au graphe précédent une application  $\Gamma$  des sommets du graphe dans l'ensemble produit des variables :



### c) Composition des lexis et des énoncés élémentaires

L'apparition de plusieurs énoncés élémentaires dans une même phrase impose une notation de la composition de ces énoncés. Le langage pivot fournit des relations destinées à remplir cette fonction. On peut s'interroger sur le nombre minimum de telles relations. Le langage pivot actuel en comporte certainement plus qu'il est nécessaire et introduit une certaine redondance.

Pour simplifier l'exposé, admettons l'existence d'une seule relation, nommée "EPITHETE". C'est un métaprédicat à deux places d'arguments :

EPITHETE(x,y)

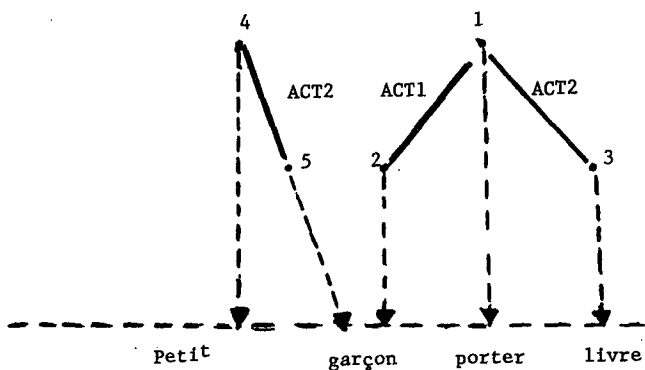
Soit la phrase : "Le petit garçon porte un livre".  
Elle contient les 2 énoncés élémentaires :

$E_1$  : Le garçon porte un livre  
provenant de la lexis : Porter(garçon, livre)

$E_2$  : Le garçon est petit  
provenant de la lexis : Petit (garçon)

Ces lexis sont représentées par les graphes :

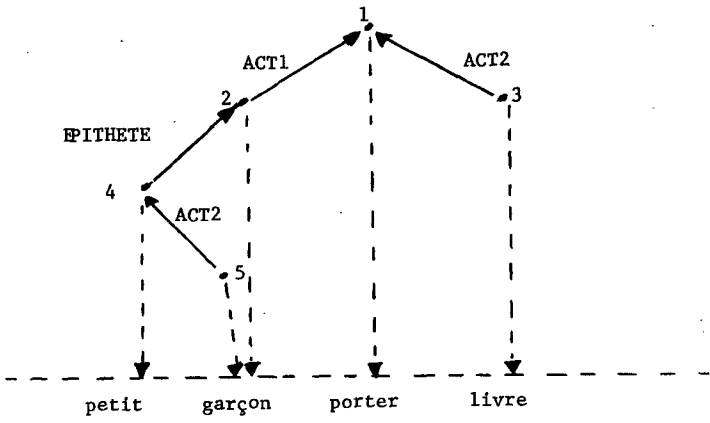




Entre les énoncés  $E_1$  et  $E_2$ , la phrase donnée en exemple implique la relation EPITHETE<sup>1</sup> entre les points 4 et 2 du graphe ci-dessus :

EPITHETE [Petit [ACT1(garçon, Porter(x,livre))],  
ACT1(garçon,porter(x,livre))]

ce qui conduit à la notation sur les lexis :



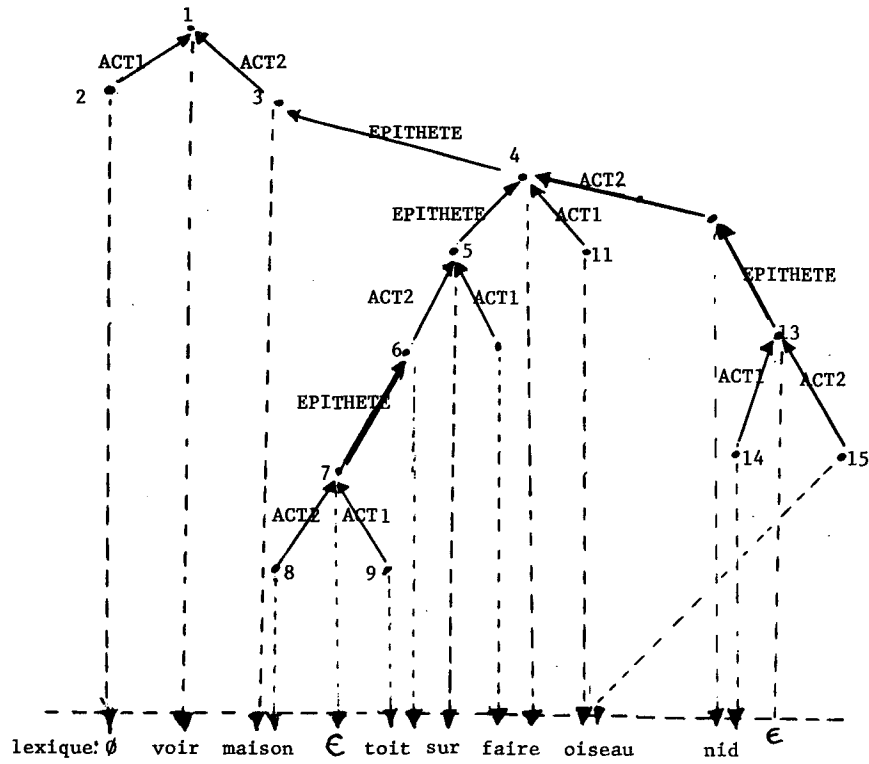
Mettons en évidence la structure des lexis sur un exemple plus compliqué

"On voit la maison sur le toit de laquelle les oiseaux font leur nid"

Les énoncés élémentaires proviennent des lexis suivantes :

- VOIR ( $\emptyset$ , maison)
- FAIRE (oiseau, nid)
- SUR (faire, toit)
- $\in$  (Toit, maison)
- $\in$  (nid, oiseau)

dont le graphe des relations est :



- III- ACCES AUTOMATIQUE AU LANGAGE PIVOT

a) Structure syntaxique initiale

La structure syntaxique initiale est obtenue à la suite de l'application des modèles précédents (M1-modèle morphologique ; M2-modèle syntaxique). Elle se présente sous forme de graphe de dépendances où chaque sommet est caractérisé par un numéro d'occurrence (numéro d'ordre dans le texte), un syntagme élémentaire ou bas comprenant la catégorie terminale et le numéro de dérivation, un syntagme non élémentaire ou haut, comprenant des valeurs de variables grammaticales et la catégorie non terminale et enfin, l'adresse dans le dictionnaire ou le numéro d'unité lexicale. L'arc reliant un sommet à son gouverneur comporte, en outre, le numéro de la règle syntaxique appliquée. (Voir, en annexe page I, l'exemple).

Le contenu d'une structure syntaxique

Etant purement formelle, une structure syntaxique comporte le plus souvent plusieurs interprétations possibles de relations logiques existant entre les sommets de son graphe. C'est ainsi que celle qui est obtenue par la relation V201 (voir l'exemple en annexe page I) implique différentes relations de prédicat à argument, selon les valeurs de variables associées au prédicat ainsi que la classe dite "sémantique" à laquelle celui-ci appartient. Si le prédicat est à la forme passive ou réfléchie transitive imperfective, le sommet relié à lui par la relation en question sera interprété comme son deuxième argument. Autrement, il le sera le plus souvent comme son premier argument.

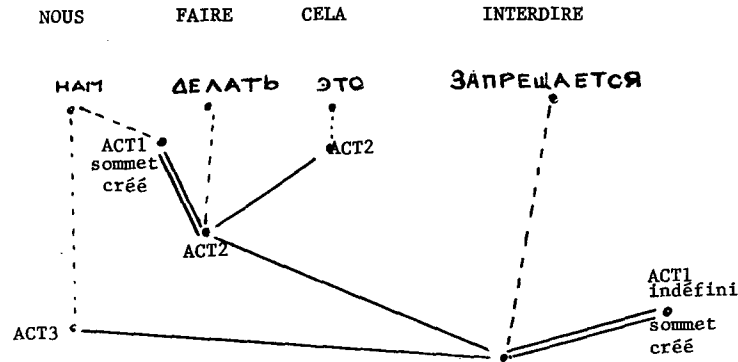
D'autre part, une structure syntaxique peut, en réalité en impliquer plusieurs autres, le choix de la meilleure parmi celles-ci ne s'étant pas révélé possible. C'est le cas, par exemple, du rattachement des compléments ou circonstants prépositionnels, conjonctifs ou adverbiaux à leur gouverneur. Etant donné qu'au niveau de l'analyse syntaxique il est inefficace et parfois impossible d'introduire des critères multiples permettant de retrouver le vrai gouverneur d'un tel sommet, nous avons opté pour une solution unique, celle de rattacher systématiquement tous les sommets de ce type au gouverneur le plus haut de la structure (voir exemple en annexe page I)

Cela implique les objectifs suivants du langage pivot :

trouver la relation exacte existant entre les sommets d'un graphe  
parmi une famille de structures contenues dans une seule structure  
donnée, trouver une structure correcte.

Il en résulte des modifications de structure à la suite de la disparition d'occurrences-outils (exemple ci-dessous), de la création de sommets supplémentaires pour faire intervenir certaines occurrences dans la structure autant de fois qu'elles apparaissent comme arguments absents de prédicats (substitution et élision) ou bien pour pallier l'omission par l'expression de certains arguments sous-entendus mais n'ayant pas de références à d'autres occurrences (indéfini, impersonnel etc...)

Exemple :



b) Description de la grammaire d'étiquetage (M3)

Pour effectuer toutes ces transformations nous recourons à une grammaire dite "d'étiquetage" dont le métalangage est décrit dans le document (8).

Cette grammaire comporte quelque 80 règles. Une règle comprend : un numéro, la partie gauche et la partie droite.

la partie gauche consiste en une description du graphe faite à l'aide des symboles de relations décrites dans (8)  
 Les sommets du graphe ainsi notés sont numérotés de 0 à n dans l'ordre de présentation

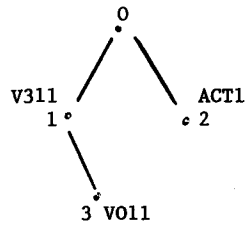
Exemple :

GAUCHE ((FS, V311 ((FR,ACT1)(FS,V011))))

EXEMPLE DE REGLE D'ETIQUETAGE

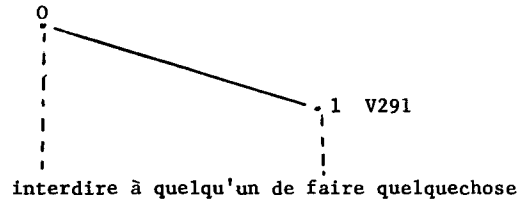
	REGLE	R220,SUITE	GIQ25910	2590
	GAUCHE	((FS,V291))	GIQ25920	2591
	TABLO		GIQ25930	2592
A	LIGNE	(CE,TRIVA)	GIQ25940	2593
B	LIGNE	(R(FS(B1,.L)))	GIQ25950	2594
B1	LIGNE	((E,V132)U(E,V151))	GIQ25960	2595
C	LIGNE	(R(FS,ACT2))	GIQ25970	2596
D	LIGNE	(R(FS,ACT3))	GIQ25980	2597
	FTABLO		GIQ25990	2598
	CONDIT		GIQ26000	2599
	STRUCT	((1,BENJ,2))	GIQ26010	2600
	SYMBO	2((E,V201))	GIQ26020	2601
	SINON	(B,0)AA	GIQ26030	2602
	STRUCT	((0,BENJ,3))	GIQ26040	2603
AA	SYMBO	2((SBT,..,3))	GIQ26050	2604
	SINON	((C,0)U(D,0))ELIMIN	GIQ26060	2605
	SINON	(A,0)AB	GIQ26070	2606
	SYMBO	1((E,ACT3))	GIQ26080	2607
	SYMBO	3((E,ACT2))(A,SUITE)	GIQ26090	2608
AB	SYMBO	1((E,ACT2))	GIQ26100	2609
	SYMBO	3((E,ACT3))	GIQ26110	2610
	FREGLE	R220	GIQ26120	2611

Dans une structure, on recherche un graphe du type

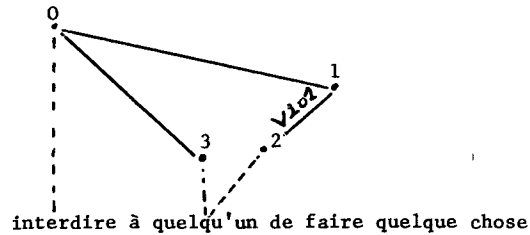


La partie droite consiste en une attribution de nouveaux symboles aux sommets définis et en une transformation éventuelle du graphe

Dans l'exemple ci-dessus, la partie gauche demande un graphe du type



La partie droite (CONDIT) le transforme immédiatement, en créant le sommet 2 et en cherchant le sommet 3, s'il existe, ou en en créant un, s'il n'existe pas. Le sommet 2 créé est muni d'une étiquette V201 et d'une adresse qui est celle de 3



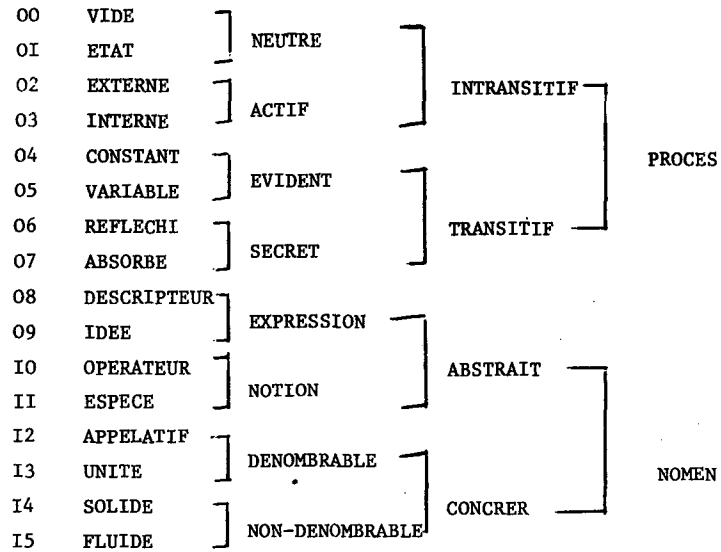
Ensuite, les sommets 1 et 3 sont munis d'étiquettes ACT2, ACT3 en fonction des codes du sommet 0. En effet, si le 0 est du type de "interdire", le 3 sera son ACT3 et le 1 son ACT2 ; si le 0 est du type de "prier", les étiquettes de 1 et 3 seront inversées.

Les codes étiquetage utilisés dans l'exemple donné ci-dessus, représentent les notations de classes dites sémantiques auxquelles appartient tel ou tel mot. Ces classes sont réunies selon des critères qui diffèrent d'une langue à l'autre. Pour le russe, nous suggérons trois critères principaux :

1) transitivité ; 2) valence ; 3) gouvernement de prépositions.

C'est en fonction de ces critères que nous avons établi trois classifications du vocabulaire, les deux premières sous forme d'arbres binaires, la troisième sous forme de liste.

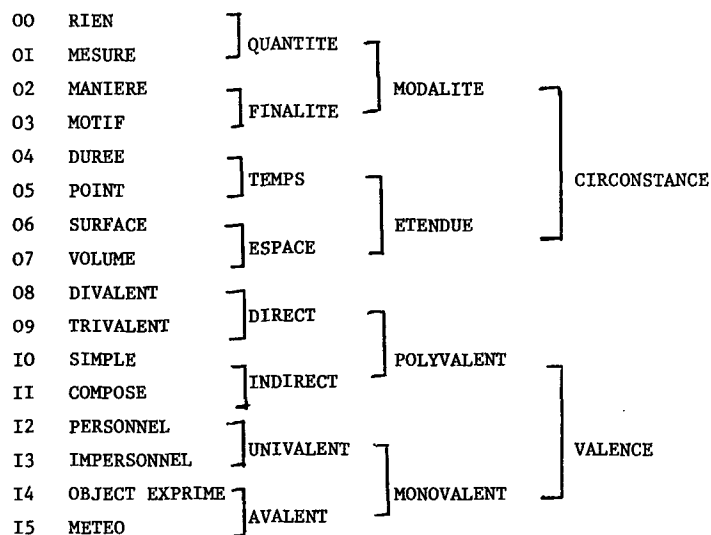
ARBRE N°1



ATTRIBUTION DES VALEURS

- 00 mots-outils, prépositions, pronoms, adverbess simples
- 01 qualificatifs dont le sujet est ACT2
- 02 action intransitive dont le sujet est ACT2
- 03 action intransitive dont le sujet est ACT1
- 04 action transitive dont le sujet est ACT1 si NRF et ACT2 si REF
- 05 action transitive dont le sujet est ACT1E2
- 06 action transitive dont le sujet est ACT1 si NRF et ACT1E2 si REF
- 07 action transitive dont le sujet est ACT1 pour REF ou NRF





## ATTRIBUTION DES VALEURS

CIRCONSTANCE : Porte sur les prépositions, adverbes, adjectifs et noms non verbaux

VALENCE : porte sur les verbes et noms verbaux

- 08 Verbe à 2 actants : ACT1, ACT2
- 09 verbe à 3 actants : ACT1, ACT2, et ACT3 (exprimé par DATIF)
- 10 verbe à 3 actants : ACT1, ACT2(au dat), ACT3 (pas DATIF)
- 11 verbe à 3 actants : ACT1, ACT2(à ACC), ACT3 (pas DATIF)
- 12 verbe à 1 actant : ACT1
- 13 verbe à 2 actants : ACT1, ACT2 mais pouvant avoir l'Attribut de l'objet
- 14 verbes impersonnels du type : il faut, possible
- 15 verbes météorologiques du type : il pleut

Il se présente sous forme suivante :

Forme canonique source	N° identificateur			I.G. Informations grammaticales	C.E. codes d'étiquetage	Champ sémantique	Equivalent	U.S. Unité sémantique	P.M.
	U.L. Unité lexicale	C.C. Code classe	D dérivation						

La forme canonique source et l'équivalent sont notés par commodité de maniement

Seuls UL, CC, D constituent le numéro identificateur du mot

Les I.G. représentent un groupe de valeurs de variables capable de particulariser telle ou telle forme du mot pour lui associer un équivalent spécial.

Les C.E. sont notés par une succession de symboles (chiffres ou lettres) séparés l'un de l'autre par une virgule et indiquant la valeur associée au mot dans les arbres d'étiquetage.

Le champ sémantique est un code de microglossaire particularisant l'emploi du mot avec un équivalent donné dans tel ou tel domaine du langage.

L'U.S. est un numéro d'ordre d'en ensemble de mots appartenant à une même famille au sens général de ce mot.

Une famille sémantique peut contenir des mots à fonctions significatives suivantes :

- 01 : Procès
- 02 : Chose
- 03 : Qualificatif actif
- 04 : façon
- 05 : caractère
- 06 : qualificatif passif.

Exemple de familles

(Les positions vides, en cas de nécessité, peuvent être remplies à l'aide de périphrases)

exemples	PM	1	2	3	4	5	6
PROCES	01	distinguer	comprendre				
	11						
CHOSE	02	distinction	compréhension				Absence
	12		incompréhension				
QUALIFICATIF ACTIF	03	distinct	compréhensif				Absent
	13		incompréhensif				
FACON	04	distinctement		réellement	pour	ET	
	14	indistinctement					
CARACTERE	05			réalité			
	15			irréalité			
QUALIFICATIF PASSIF	06	distingué	compris	REEL	pour que		
	16		incompris	irréel			-17

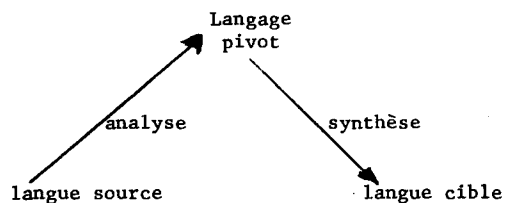
Ce langage de description a été conçu dans le cadre d'un programme de traduction automatique et nous traiterons plus particulièrement de cette application.

a) Utilisation dans un programme de traduction automatique

Cette notation préserve la signification en éliminant les contraintes d'expression syntaxique et morphologique. Ceci correspond effectivement à l'indépendance recherchée dans un modèle de type "sens-texte".

Nous ne prétendons pas atteindre effectivement une notation de la signification, mais plus simplement, nous recherchons une structure indépendante des langues. Ces dernières étant caractérisées par les contraintes d'expression que nous avons éliminées.

Dans son utilisation pratique, ce langage intermédiaire a été simplifié. En particulier, certaines analogies entre les langues permettent d'éviter une analyse trop profonde et de lever certaines ambiguïtés. Dans le système de traduction automatique, cette notation constitue le pivot à partir duquel il est possible de produire une expression en langue cible. Ainsi, analyse et synthèse sont complètement distinctes, selon le schéma suivant :



Les procédés d'analyse ayant été déjà commentés, nous étudierons plus particulièrement la synthèse.

b) Nature du lexique

La notation en langage pivot doit permettre de produire une structure syntaxique de surface en langue cible, dont on déduira ensuite la chaîne de mots. Nous supposons que la langue cible est le Français. A chaque sommet de la structure est associée une unité du lexique, déterminée à partir de l'analyse de la langue source. Dans le système de production, chaque unité du lexique est caractérisée par plusieurs expressions de propriétés syntaxiques différentes.

Elles correspondent aux douze possibilités définissant un paradigme étendu (voir figures et exemples).

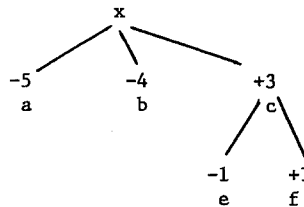
A chacune de ces expressions sont associées des informations grammaticales spécifiant le comportement syntaxique du mot : catégorie syntaxique, genre, nombre, prépositions gouvernées, la racine et les propriétés morphologiques.

Un des rôles de la grammaire de production est de choisir la catégorie syntaxique de chacune des unités en fonction des contraintes de la structure et de la grammaire.

c) Notation de la structure syntaxique

Le graphe sera obtenu par transformations successives du graphe représentant la notation en langage pivot. Nous obtiendrons donc une structure arborescente du type de dépendance. Cependant, nous n'utiliserons pas de règles de productions au sens des grammaires génératives. De ce fait, la relation d'ordre totale entre les mots n'est pas exprimée par la structure autrement que par l'habituelle contrainte de projectivité. Si l'on tient compte de celle-ci, il suffit d'indiquer pour chaque sommet sa position vis à vis du sommet qui le gouverne. Cette position sera spécifiée par un nombre entier positif si le sommet est à droite, négatif si le sommet est à gauche du gouverneur.

Exemple

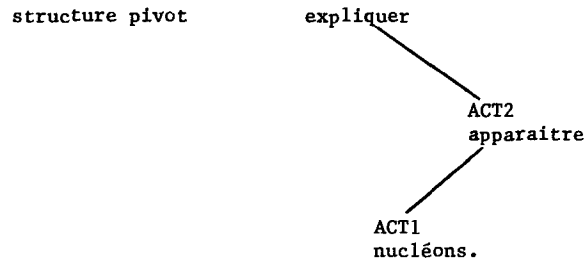


Les sommets a, b et c sont ordonnés par rapport à x : a b x c  
et e et f : e c f  
La règle de projectivité définit la chaîne : (a b x (e c f)).

La structure syntaxique de surface se présentera donc comme un graphe arborescent dans lequel chaque sommet sera marqué par : un 'poids' (nombre entier caractérisant sa position), un nom de catégorie syntaxique et un ensemble de variables grammaticales permettant de construire la forme (voir exemple annexe page 3).

d) La grammaire de production

Pour chaque sommet, deux types d'information doivent être déterminés : la fonction syntaxique et la catégorie syntaxique. Nous ne pouvons déterminer la catégorie qu'en connaissant la fonction, ce que montre l'exemple suivant :



Supposons que nous ayons imposé la catégorie nominale à la racine, exprimée par "explication". Dans ce cas, la fonction syntaxique de "apparaître" sera une fonction de complément de nom (génitif) qui nous impose alors une catégorie nominale : "explication de l'apparition de nucléons".

Si nous avons une catégorie verbale "expliquer", alors la fonction syntaxique de "apparaître" sera une fonction de complément d'objet, exprimable soit par une catégorie nominale, soit par une catégorie verbale (expliquer que ...).

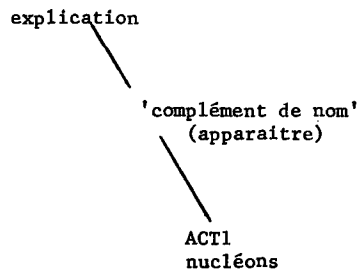
La grammaire de production est appliquée sur chaque sommet, en commençant par la racine. Pour chaque sommet, trois types successifs de règles sont appliqués :

1) Choix de la catégorie syntaxique : connaissant la fonction syntaxique du sommet, la détermination de la catégorie syntaxique dépend des possibilités d'expression de l'unité lexicale, et éventuellement de son environnement

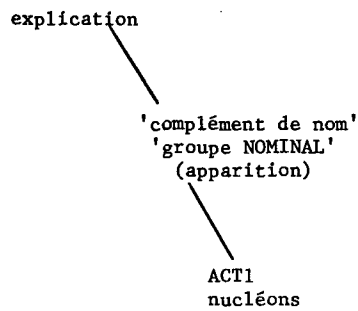
2) Détermination des fonctions syntaxiques des divers dépendants du sommet sur le graphe pivot

3) Règle terminale : Cette règle permet de déterminer la forme à exprimer en morphologie. Elle peut effectuer des transformations dans certains cas : expression du temps composé, des prépositions et des articles, transformation de la structure au cas où une catégorie nominale correspond à un pronom.

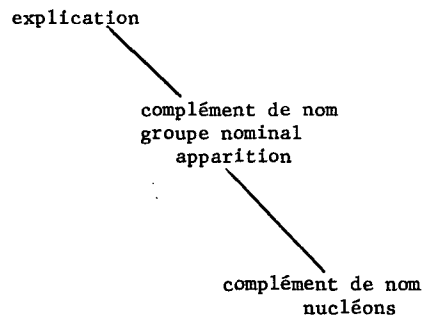
Exemple : Soit le sommet "Apparaître" de l'exemple ci-dessus. -21  
Avant l'application de la grammaire sur ce sommet, nous avons la structure suivante :



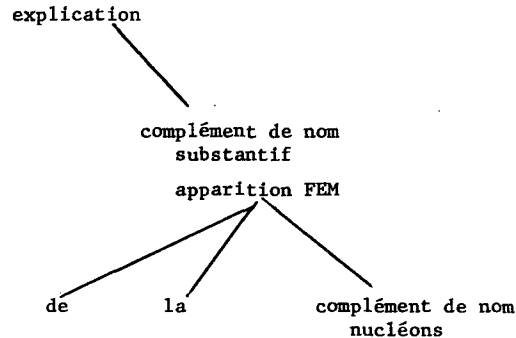
application de la règle de choix :



application des règles de définition des fonctions



Application de la règle terminale



En pratique, nous avons confondu la notion de fonction syntaxique et la notion de poids, cette dernière étant plus précise.

Autres applications

On peut envisager l'utilisation de cette notation et des systèmes d'analyse et de production qui lui sont associés, dans des systèmes de communication en langue naturelle avec un ordinateur. La notation contient le sens de plusieurs phrases équivalentes. De ce fait, on peut la considérer comme un représentant unique, forme canonique de toutes les phrases de même signification. Si l'on sait interpréter le contenu de cette notation, l'on pourra admettre dans le langage de communication homme-machine toutes les phrases équivalentes. De plus, dans un système de ce type, les possibilités de paraphrasage par production d'une ou plusieurs phrases permettent de concevoir un dialogue entre la machine et l'utilisateur. En particulier, dans le cas où une phrase proposée est ambiguë et fournit plusieurs notations, il est possible d'avoir recours à une question complémentaire afin de lever l'ambiguïté.

Parmi les applications de ce type, nous retiendrons particulièrement l'interrogation de fichiers et l'enseignement programmé.

Cependant, on doit aussi considérer l'interprétation de cette notation qui a été conçue en vue des problèmes de traduction, dans le cadre de ces nouvelles applications. En particulier, dans le cas de l'interrogation de fichiers, il est nécessaire de produire à partir de la notation, une certaine séquence de commandes, voire même un véritable programme.



## STRUCTURE SYNTAXIQUE RUSS

000070

0000\*\*\*'DEBUT DE PHRASE'

```

*
*
*           ND61***ATMOSFJERNYJ
*           *
*           *
*           V201***VOZDUKH
*           *
*           *
*           V681***,
*           *
*           *
*           P160***SZHIMAJAS'
*           *
*           *
*           *           V570***ZA
*           *           *
*           *           *
*           *           *           M271***SCHJET
*           *           *
*           *           *           *
*           *           *           *           ND61***SKOROSTNOGO
*           *           *           *           *
*           *           *           *           *
*           *           *           *           *           N124***NAPORA
*           *           *
*           *           P140***,
*           *
*           *
*           H021***NAGRJEVAJETSJA
*           *
*           *
*           *           V570***V
*           *           *
*           *           *
*           *           *           M271***TJEPLOOBMJENNIKJE
*           *
*           *
*           H180***.

```

```
) *DEBUT DE PHRASE *
*
*   ACT2 *VOZDUKH *AIR
*
*   PHRASE*NAGRJEVAJETSJA CHAUFFER
*
*   CIRGEN*V DANS
*
*   ACT3 *TJEPLOBMJENNIKJE *ECHANGEUR
*
*   GEROND*SZHIMAJAS' COMPRIMER
*
*   CIRGEN*SCHJET GRACE
*
*   ACT2 *SUBSTITUE'
*
*   EPITHE*SKOROSTNOGO DYNAMIQUE
*
*   ACT3 *NAPORA *EXERCER
*
*   ACT1E2'SUBSTITUE'
*
PUNCTU*.
```

** +0 *DEBUT DE PHRASE*		VIT
GNOMO** -50 *AIR		VIFVICVIOVIAMANI NDMASSINTREINTARDVIT
O(CB/ * -16 6LE1		VIFVICVIDVIAMASSINVIT
VERBO*** +1 *EST		VIFVICVIOSSGAVRSA XVI AI NDPREMASS INTREVPPVIT
ADVB** +85 *DANS		VIFVICVIOSSCADGV IA INDIAMASSININTVIT
GNOMO*** +1 *ECHANGEUR		VIFVICVICVIAMANI NDMASSINTREINTARDVIT
O(CB/ * -16 6LE1		VIFVICVIDVIAMASSINVIT
GNOMO** +18 *CHALEUR		VIFVICVICVI AFMPS MIT RMFEMS INTNRARRDET
O(CB/ * -18 6E8		VIFVICVIOSSCADGVIAVIT
VERBO** -52 *COMPRIMAINT		VIFVICVIOSSGETRRFLVIA INDPREMASSINTREVIT
O(CB/ * -45 5EN		VIFVICVIOSSCADEVIAVIT
O(CB/ * -40 5E1		VIFVICVIDVIAMASSINTREREFVIT
ADVB** +85 *GRACE		VIFAACVIDSSCADGVIA INDI NAMASSI INTREINTVIT
GNOMO*** +1 *PRESSION		VIFVICVIDVIAFM INDFEMSINTREINTARJAAIT
O(CB/ * -16 6L1		VIFVICVIDVIAFEMS INVIT
O(CB/ * -18 5A9		VIFVICVIDSSCADGVIAVIT
ADJF*** +12 *DYNAMIQUE		VIFVICVICVI IA INDPREFEMS INTREVIT
O(CB/ +100 ,		VIT
ADJF*** +46 *CHAUFFE		VIFVICVICSSGAVRVIA INDPASSINTREVIT
PUNCTO +100 .		VIT

#### BIBLIOGRAPHIE

- (1) - A. JOLKOVSKY - Y. MELTCHOUK  
Essai d'une théorie sémantique applicable au traitement de langage - Annales de la Conférence Internationale sur le traitement automatique des langues - Grenoble 1967
- (2) - А. ЖОЛКОВСКИЙ , И. МЕЛЬЧУК  
О семантическом синтезе  
Проблемы Кибернетики - 19-1967 - 177-228
- (3) - S. LAMB  
Outline of stratificational grammar  
Georgetown University Press - 1966
- (4) - B. VAUQUOIS  
Le système de traduction automatique du C.E.T.A.  
Congrès d'EREVAN - Avril 1967
- (5) - G. VEILLON  
Description du langage pivot du système de Traduction automatique du C.E.T.A.
- (6) - C. FUCHS - M. PECHEUX  
Lexis et Métalexis  
"Linguistique mathématique" - Collection de D. HERAULT (Dunod)  
(à paraître)
- (7) - M. DUPRAZ - J. ROUAULT  
Lexis - Affirmation - Négation  
Etude fondée sur les Classes  
Colloque de Balatonszabadi - Septembre 1968
- (8) B. VAUQUOIS - G. VEILLON  
Un métalangage de grammaires transformationnelles  
Proceeding of International Conference on Computational Linguistics - Grenoble 1967.