

BERNARD VAUQUOIS

**Modèles pour la traduction automatique**

*Mathématiques et sciences humaines*, tome 34 (1971), p. 61-70

[http://www.numdam.org/item?id=MSH\\_1971\\_\\_34\\_\\_61\\_0](http://www.numdam.org/item?id=MSH_1971__34__61_0)

© Centre d'analyse et de mathématiques sociales de l'EHESS, 1971, tous droits réservés.

L'accès aux archives de la revue « Mathématiques et sciences humaines » (<http://msh.revues.org/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## MODÈLES POUR LA TRADUCTION AUTOMATIQUE

par

Bernard VAUQUOIS <sup>1</sup>

*Lorsqu'on s'efforce de poser le problème à un niveau suffisamment abstrait et de le formuler en termes théoriques, la traduction automatique a été et reste une motivation importante pour une étude opérationnelle des langues naturelles. En vue d'expliquer le « comment » des processus de production et de reconnaissance d'une langue (c'est le cas de la traduction), on peut concevoir un automate qui produise ou reconnaisse un texte en effectuant des calculs effectifs au sens de calculs effectués par une Machine de Turing, mais on peut exiger en plus que tout calcul effectué par cet automate soit effectué en un temps « raisonnable », avec une mémoire limitée, par une machine réelle. Si l'on veut tester la validité d'une théorie linguistique et en particulier d'une grammaire de production ou de reconnaissance, n'est-il pas meilleur test que le « passage » en machine réelle? L'expérimentation, à l'aide d'un ordinateur, permet de poser de nouveaux problèmes, de réaliser des combinaisons complexes, d'imposer au théoricien des contraintes qui éviteront parfois les solutions évidentes!*

J. P. DESCLÉS

### INTRODUCTION

La traduction, du moins telle qu'elle est conçue depuis une dizaine d'années, a été une motivation importante pour l'étude et la réalisation de modèles calculables par ordinateur. Par la suite, d'autres centres d'intérêt tels que la communication homme-machine en langue naturelle ou encore les problèmes d'indexage en documentation automatique ont profité de la même démarche méthodologique et parfois emprunté une partie des mêmes réalisations.

Aussi, dans ce qui suit, si la plupart des exemples sont choisis pour les besoins de la traduction automatique, il ne sera pas défendu néanmoins, de montrer comment ces mêmes modèles peuvent servir à autre chose.

Ce cadre, plus large que celui de la seule traduction automatique sera désigné « traitement automatique des langues » (en abrégiation : T.A.L.).

Une traduction, automatique ou non, est soumise à un certain nombre de contraintes. En particulier, le texte à traduire est considéré du point de vue de son contenu et non de sa forme d'expression. On élimine ainsi les textes dont la forme prend une part importante (art littéraire). La « traduction » de tels textes est en réalité une nouvelle composition artistique.

---

1. CNRS, Centre d'études pour la traduction automatique.

En traduction automatique, il est nécessaire de prolonger les conséquences de la remarque précédente aussi loin que possible. On considère que la langue d'expression est le moyen de codage d'un « message » qui aurait pu être codé dans plusieurs autres langues. Par exemple, la description d'une expérience de physique peut aussi bien s'écrire en français, anglais, russe, japonais, etc.

Il est aussi nécessaire de soumettre à la machine des textes « correctement » écrits. En effet, l'idée de modèle ne peut s'appliquer qu'à des niveaux de langue bien définis et décrits rigoureusement. Il est impossible, ou du moins totalement inefficace, de demander à une machine d'interpréter correctement toutes les fantaisies de syntaxe, de lexique, d'orthographe... auxquelles se livrent certains auteurs scientifiques. Ces restrictions étant acceptées (il y en aura d'autres qui interviendront par la suite), voyons ce que représente la notion de modèle en traitement automatique des langues.

## I. NOTION DE MODÈLE ET NIVEAUX DE LANGUE

La notion de modèle qu'on désire appliquer n'a rien d'original et se trouve couramment pratiquée dans d'autres disciplines. Ce que l'on souhaite obtenir en traitement automatique des langues, ce sont des modèles adaptés aux différentes applications fixées comme objectif. Dans ce but on est conduit à distinguer divers niveaux de langue.

Le niveau le plus bas est celui du texte considéré comme chaîne de symboles (chaîne de caractères alphabétiques, ponctuation, etc., pour un texte écrit, chaîne de symboles phonétiques pour un texte parlé, préalablement transcrit).

C'est le niveau habituel où les textes sont introduits dans un ordinateur si l'on n'effectue aucune prédiction manuelle. En traduction automatique c'est le niveau de départ du texte à traduire et le niveau de sortie du texte traduit. Il en est de même à l'entrée et sortie d'un système de communication homme-machine.

On peut imaginer à l'autre extrémité un niveau de « contenu » qui représente la signification du texte codé par la chaîne de symboles. Ce niveau ne se prête à aucune définition rigoureuse et apparaît inaccessible. On se contentera donc de niveaux intermédiaires entre le « texte » et le « contenu ».

Pour fixer les idées, disons que, dans la hiérarchie ascendante texte  $\rightarrow$  contenu, un niveau  $n_2$  sera supérieur à un niveau  $n_1$  si la formulation du message dans  $n_2$  échappe davantage aux contraintes de la langue que la formulation du même message dans  $n_1$ .

Par exemple, prenons un message limité à une phrase de texte : cette phrase est transformée en une formule  $f_1$  au niveau  $n_1$  et en une formule  $f_2$  au niveau  $n_2$ . Alors le nombre de phrases de signification équivalente ou représentées par la même formule  $f_2$  est en général supérieur (quelquefois égal) au nombre de celles représentées par la même formule  $f_1$ . En somme, plus le niveau est élevé et plus la capacité de paraphrasage est grande.

A titre d'exemple simple, donnons-nous les deux phrases :

(1) « Le client paie la marchandise. »

[2] « Le client paye la marchandise. »

Au niveau du texte (chaîne de caractères alphabétiques) ces deux phrases sont différentes à cause de « paie » et « paye ».

Au niveau de représentation morphologique, elles sont identiques, car « paie » et « paye » sont représentés par l'unique formule :

verbe « payer » : troisième personne singulier, indicatif présent  
la troisième phrase :

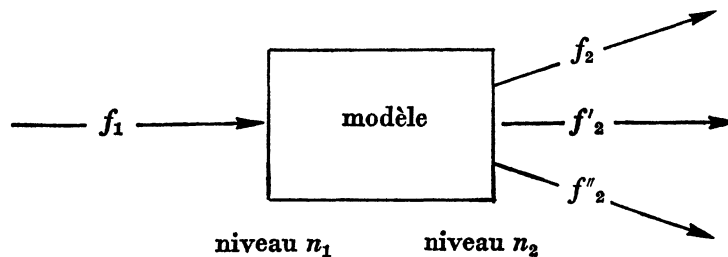
(3) « La marchandise est payée par le client »

est différente des deux premières, au niveau du texte, au niveau morphologique et aussi au niveau syntaxique. Pourtant, il est aisé d'imaginer un niveau supérieur où ces trois phrases sont représentées par la même formule. En effet, supposons que l'on note le concept de « payer » comme un prédicat à deux places d'arguments : PAYER ( $x_1, x_2$ ) où  $x_1$  est la place de l'élément payeur et  $x_2$  celle de l'élément payé, on aboutit à la formule unique :

PAYER (client, marchandise) dans les trois exemples.

Pour passer d'un niveau à un autre de manière automatique, on réalise un modèle. Il est évident que lorsque les niveaux sont déterminés, on se borne à réaliser des modèles effectuant le passage au niveau immédiatement inférieur ou supérieur. Le but de chaque modèle consiste à fournir pour chaque formule proposée au niveau d'entrée, toutes les formules correspondantes au niveau de sortie.

On dit qu'il s'agit d'un modèle d'analyse (ou de reconnaissance) si le niveau d'entrée est inférieur au niveau de sortie, qu'il s'agit d'un modèle de synthèse (ou génératif) dans le cas contraire.

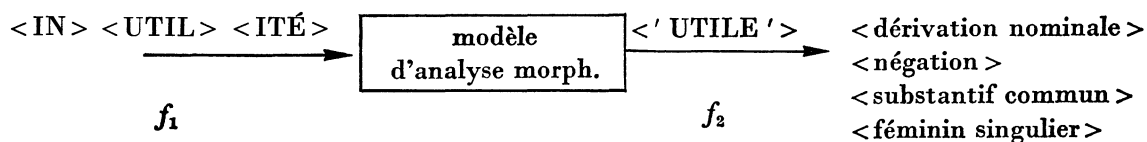


Étant donnée  $f_1$ , formule au niveau d'entrée, le modèle peut fournir :

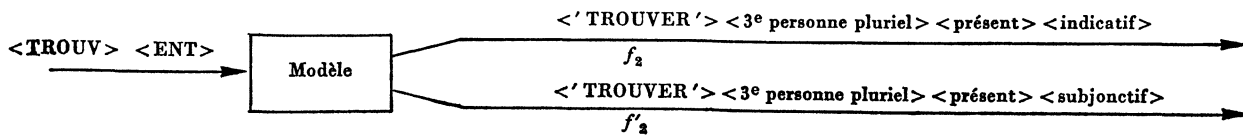
- a) Aucune formule de sortie : cela signifie que la formule  $f_1$  n'était pas valable ;
- b) Plusieurs formules de sortie : pour un modèle de synthèse cela signifie qu'il calcule toutes les paraphrases possibles ; pour un modèle d'analyse, cela signifie que la formule  $f_1$  est ambiguë.
- c) Une seule formule de sortie : pour un modèle de synthèse cela correspond à un choix (selon un critère quelconque) d'une seule solution parmi les paraphrases ; pour un modèle d'analyse cela signifie que  $f_1$  n'est pas ambiguë.

A titre d'exemple, considérons un modèle d'analyse morphologique : dans ce cas, une formule d'entrée est une chaîne de morphèmes provenant de la segmentation d'un mot (segmentation obtenue au moyen d'un dictionnaire de morphèmes) ; une formule de sortie est une chaîne de symboles qui détermine l'origine lexicale du mot ainsi que les diverses dérivations qu'il a subies et les valeurs des variables grammaticales qui lui sont attachées.

Ainsi « INUTILITÉ » est segmenté en < IN > < UTIL > < ITÉ > : on obtient :



ou encore TROUVENT segmenté en < TROUV > < ENT >.



Pour mettre en évidence un cas où aucune formule de sortie n'apparaît, il suffit de choisir un mot à segmentations multiples tel que :

## PORTES

Ce mot conduit à trois formules d'entrée :

- 1) < PORTE > < S > où < PORTE > est le substantif et < S > la marque du pluriel.

Cette formule est valable et donne un résultat.

- 2) < PORT > < ES > où < PORT > est la base du verbe porter et < ES > une désinence.

C'est aussi valable.

- 3) < PORT > < ES > où < PORT > est le substantif et < ES > une désinence.

Pour cette formule d'entrée, il n'y a pas de sortie car il s'agit d'une segmentation incorrecte.

Cet exemple de modèle morphologique permet aussi d'éclairer ce qui suit concernant les étapes à réaliser pour la construction d'un modèle.

1. On doit choisir les niveaux d'entrée et de sortie et en donner une description formalisée. Ceci signifie que l'on détermine les classes d'objets mathématiques auxquelles appartiennent les formules d'entrée et de sortie, ainsi que les contraintes à imposer à ces formules.

Ainsi, dans le cas du modèle morphologique on procède aux opérations suivantes :

- a) On définit une liste de morphèmes tels que chaque forme de mot de langue soit une suite de certains de ces morphèmes. On ne considère comme acceptable que les suites qui constituent une forme de mot et une seule. Ainsi l'unité de traitement est la forme de mot (choix du niveau d'entrée).
- b) On définit des ensembles de catégories syntaxiques, variables grammaticales, dérivations, références lexicales, etc., une formule de sortie est une suite de tels symboles (choix du niveau de sortie).
- c) La description formalisée comporte la définition de classes de morphèmes, où chaque classe est représentée par un symbole. L'ensemble de ces symboles constitue l'alphabet d'entrée. On procède de manière analogue pour créer un alphabet de sortie.

Les objets mathématiques qui expriment les formules d'entrée et de sortie sont donc des chaînes construites respectivement sur les deux alphabets (éléments du monoïde libre engendré par l'alphabet correspondant).

Dans le cas d'un modèle d'analyse syntaxique (au niveau syntaxe de surface) les formules de sorties ne sont plus exprimées par des chaînes mais par des arborescences qui sont évidemment d'autres objets mathématiques.

2. On doit ensuite, compte tenu des résultats de la première étape, ajouter aux contraintes déjà explicitées pour les formules des niveaux d'entrée et sortie, d'autres types de contraintes liées d'une part aux phénomènes linguistiques que l'on veut reproduire dans le modèle et d'autre part, aux possibilités de calcul d'une formule du second niveau à partir d'une formule du premier.

Ceci conduit à examiner, parallèlement l'adéquation du modèle à la réalité et le choix d'un *type logique*. Ce type logique est un modèle abstrait générateur de tous les modèles concrets ayant en commun les objets mathématiques qui expriment les formules d'entrée, ceux qui expriment les formules de sortie et les contraintes de calcul.

Reprenons l'exemple d'un modèle d'analyse morphologique du français. Les formules d'entrée sont exprimées par des chaînes construites sur un alphabet  $V_E$  dont les symboles sont les classes morphémiques du français selon un procédé extérieur de classification.

Les formules de sortie sont exprimées par des chaînes construites sur un alphabet  $V_S$  dont les symboles traduisent des concepts plus abstraits (classe syntaxique, nombre, temps, personne, etc.).

Ainsi le modèle doit réaliser un calcul de chaîne à chaîne. En français, il n'est pas difficile de déterminer une classification des morphèmes telle que toute forme de mot soit représentée par une chaîne de symboles obéissant à la contrainte suivante :

La validité ou le refus d'un symbole dans la chaîne ne dépend que du symbole précédent (éventuellement d'un nombre fini de symboles précédents). Dans ces conditions, le type logique de modèle le plus simple et le plus efficace est le type « transducteur à états finis ».

Ainsi ce type logique de modèle convient pour tout modèle chaîne  $\rightarrow$  chaîne dont les contraintes sont dites « d'états finis ». Il convient en particulier pour la réalisation du modèle concret de morphologie du français mais aussi pour des modèles de morphologie de beaucoup d'autres langues.

Pour un modèle d'analyse syntaxique (au niveau syntaxe de surface) les formules d'entrée seront les formules de sortie du modèle d'analyse morphologique, donc des chaînes de symboles. Par contre, que l'on se propose d'obtenir une structure syntaxique du type P-marker de Chomsky ou bien un graphe de dépendance du type Tesnières-Hays, les formules de sortie sont des arborescences. Il s'agit d'un modèle chaîne  $\rightarrow$  arborescence. Si les contraintes sur les chaînes d'entrée, contraintes imposées par les phénomènes linguistiques, sont relativement simples (continuité des constituants immédiats ou projectivité du graphe de dépendance) alors le type logique de modèle le plus simple et le plus efficace est le type dit « hors contexte » (context-free).

Il arrive déjà à ce niveau, et encore davantage aux niveaux supérieurs, que le choix d'un type logique connu et relativement simple ne soit plus compatible avec les conditions d'adéquation du modèle à la réalité. Il est donc nécessaire soit de trouver d'autres types logiques (éventuellement d'en construire spécialement) pour satisfaire pleinement aux conditions d'adéquation, soit de tronquer la langue (abandonner la représentation complète du niveau de langue considéré) pour garder un type logique simple, soit enfin de trouver une solution de compromis.

Cette étude est sans doute l'une des plus difficiles et des plus importantes du traitement automatique des langues.

3. L'étape suivante consiste à réaliser le modèle concret sur la base du type logique adopté et à le rendre opérationnel.

Parallèlement, on doit d'une part écrire une grammaire, c'est-à-dire un ensemble de règles qui permettent de calculer pas à pas, la formule de sortie à partir de la formule d'entrée, d'autre part, étudier un algorithme qui est le mode d'emploi de la grammaire ou si l'on veut le fil directeur de la conduite du calcul. Il peut, évidemment, exister plusieurs algorithmes qui utilisent la même grammaire et conduisent au même résultat. Il s'agit là d'une étude d'efficacité.

En ce qui concerne l'écriture de la grammaire, il est nécessaire de disposer d'un formalisme d'écriture (métalangage dans lequel la grammaire est écrite). Chaque type logique fournit un tel forma-

lisme (quelquefois même plusieurs). Tel est le cas, par exemple, de la forme normale de Chomsky ou celle de S. Greibach à propos des grammaires hors contexte.

Cependant, l'expérience montre que ces formalismes, bien adaptés à l'étude des propriétés intrinsèques des modèles, se prêtent mal à une utilisation en vue de l'écriture d'une grammaire réelle. En effet, l'utilisation directe du formalisme de Chomsky, par exemple, conduirait à l'écriture de plusieurs dizaines de milliers de règles pour un modèle d'analyse syntaxique du russe, du français, etc.

Il est donc indispensable de déterminer des métalangages, équivalents aux formalismes utilisés dans la théorie, qui soient opérationnels.

A l'aide de ces outils le modèle concret peut enfin apparaître sous la forme d'une première version.

4. Enfin, la dernière étape consiste à vérifier l'adéquation effective du modèle. Il s'agit là de corriger les erreurs dans les règles de grammaires ainsi que celles qui proviennent du codage (affectation à des classes) des éléments lexicaux. Ceci représente un travail analogue à celui qui consiste à mettre au point un programme. Cependant, la « mise au point » d'une grammaire et d'un dictionnaire est une opération beaucoup plus difficile pour plusieurs raisons :

— En effet, il n'est pas toujours possible de segmenter une grammaire comme on peut segmenter un programme. En outre, on ne dispose d'aucune méthodologie rigoureuse pour déceler toutes les conséquences de la correction d'une règle de grammaire.

Pour assurer l'adéquation effective du modèle à 90% il est nécessaire d'avoir contrôlé des milliers de formules d'entrées. Ainsi, un modèle d'analyse syntaxique commence à avoir une valeur réelle lorsque cinq ou six mille phrases ont été contrôlées.

Ce travail est donc très long et devient fastidieux. Pourtant, il est indispensable de le conduire jusqu'au bout. Il y a dans ce domaine, un travail de recherche difficile pour systématiser et accélérer la mise au point des modèles.

## II. APERÇU SUR QUELQUES MODÈLES ET LEURS APPLICATIONS

Il serait hors de propos de vouloir dans ces quelques lignes entrer dans le détail de chacune des étapes précédemment mentionnées pour des modèles relatifs à tous les niveaux qui ont été réalisés ou sont en cours de réalisation.

Nous nous bornerons à exposer les résultats de certains d'entre eux et à montrer quelques-unes de leurs applications.

### 1. MODÈLES D'ANALYSE MORPHOLOGIQUE

Dans la mesure où une langue possède une morphologie assez riche, c'est un modèle qui s'impose pour toute étude visant le traitement automatique des langues. Faute de quoi, on serait obligé de construire des dictionnaires gigantesques pour reconnaître toutes les formes de mots.

Ainsi l'inutilité de ce modèle apparaît dans deux cas :

a) La morphologie de la langue est très pauvre : le nombre de formes de mots n'est que très peu supérieur au nombre de racines ;

b) Le vocabulaire du sujet traité est très restreint : il n'apparaît que peu de mots et ces mots n'utilisent jamais la plupart des formes possibles (titres, éventuellement résumés succincts d'articles).

## Applications

a) Étude de vocabulaire par procédé automatique.

En effet, le même mot est reconnu sous toutes ses formes. On peut aussi étudier l'emploi des dérivations et leur complexité dans un texte quelconque.

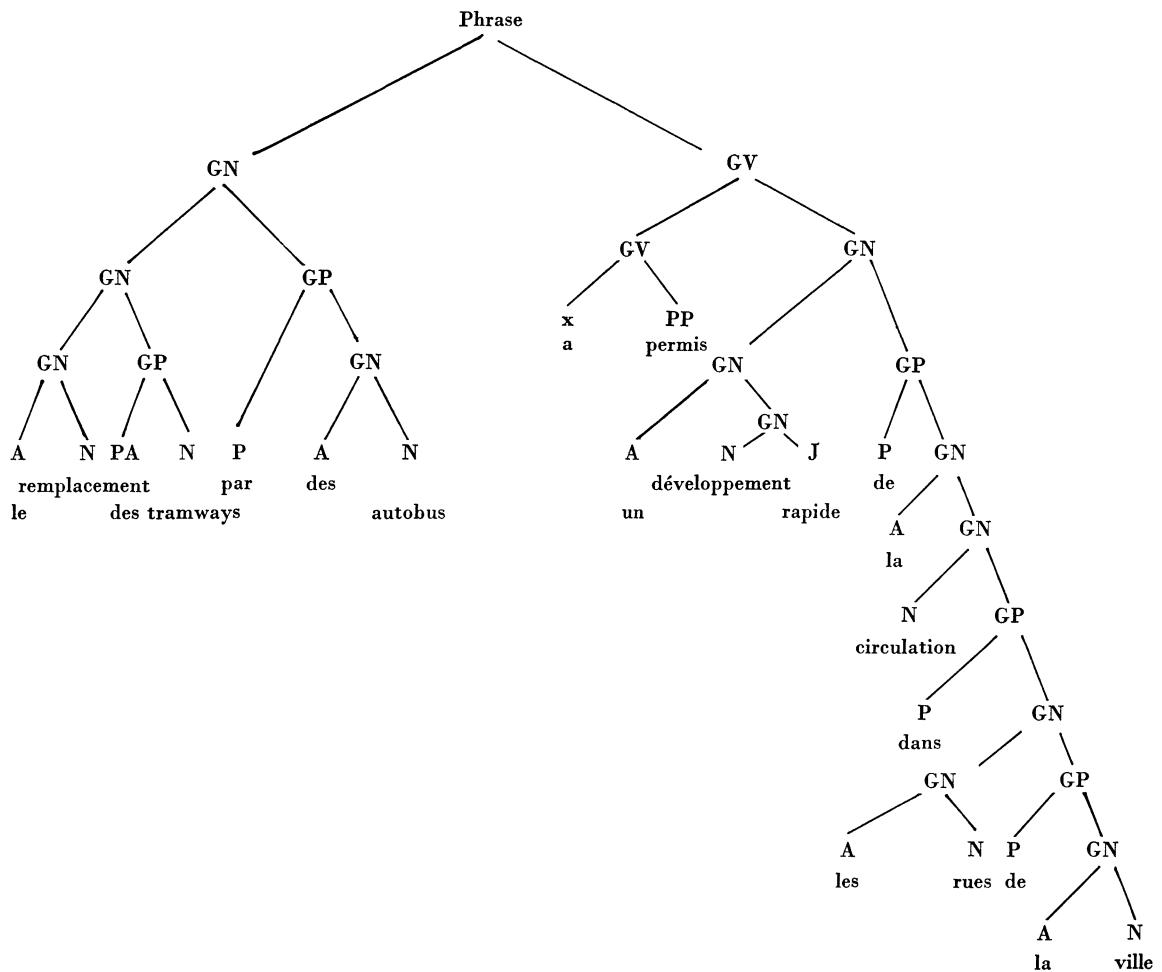
b) Étape intermédiaire pour des études qui nécessitent l'accès à un niveau de langue plus élevé.

## 2. MODÈLES D'ANALYSE SYNTAXIQUE (AU NIVEAU SYNTAXE SUPERFICIELLE)

On atteint là un niveau déjà intéressant. Avant d'énumérer les multiples applications d'un tel modèle, voyons d'abord quelle est l'allure des résultats sur un exemple, peut être un peu compliqué mais qui pourra être réutilisé à d'autres niveaux.

Soit la phrase : « Le remplacement des tramways par des autobus a permis un développement rapide de la circulation dans les rues de la ville. »

Il existe des modèles d'analyse syntaxique qui vont fournir la formule suivante au niveau de sortie.





En réalité, la formule est plus compliquée car elle comporte (ce qui n'est pas noté ici) l'identification de la règle de grammaire qui a construit chaque syntagme intermédiaire et les valeurs des variables grammaticales qui sont associées à chacun d'eux.

En outre, les mots de la phrase sont en réalité les chaînes de symboles qui résultent de l'analyse morphologique.

Il existe aussi d'autres formulations à ce même niveau, qui représentent également la construction syntaxique au moyen d'une arborescence (schémas de Tesnière-Hays, schémas obtenus à partir des grammaires d'ajouts de Joshi-Fraser-Salkoff).

Les applications que l'on trouve sont déjà nombreuses.

a) *Ce niveau a été utilisé comme niveau de transfert* dans certaines expériences de traduction automatique notamment au Japon, aux États-Unis et partiellement en Grande-Bretagne. Dans de telles expériences on établit une correspondance entre des « sous-arbres », correspondant à des groupes syntaxiques dans la langue source et dans la langue cible.

On substitue ainsi, par morceaux, des sous-arbres (éléments de formule de la structure syntaxique en langue cible) à ceux qui ont été obtenus par l'analyse syntaxique de la langue source jusqu'à la création de la structure syntaxique complète en langue cible.

Il reste alors à remplacer les mots du texte initial par leurs équivalents (appartenant à la classe syntaxique exigée par la nouvelle structure) dans la langue cible et à réaliser la construction du texte traduit au moyen d'un modèle de synthèse.

Les limites d'une telle méthode sont évidemment subordonnées aux possibilités de correspondance entre groupes syntaxiques et aussi à la capacité de résolution des ambiguïtés du modèle d'analyse.

b) *Des applications pédagogiques variées.* En effet, on peut soumettre à de tels modèles d'analyse syntaxique les textes écrits par les élèves qui ont étudié une langue pendant 1 an, 2 ans, etc., et observer ainsi la progression dans la complexité des structures syntaxiques qu'ils utilisent.

c) Des applications linguistiques visant la comparaison des structures syntaxiques de surface entre diverses langues.

d) Ce niveau est bien adapté à la synthèse automatique de la parole pour calculer la ligne mélodique (pitch), la segmentation en groupes phoniques et la durée relative des phénomènes.

On peut ainsi (des expériences sont en cours) réaliser la conversion automatique d'un texte écrit en texte parlé.

e) Enfin, ce niveau (éventuellement un niveau très voisin) est un intermédiaire inévitable pour accéder à un niveau de langue supérieur.

### 3. MODÈLES DE NIVEAU SUPÉRIEUR A LA SYNTAXE DE SURFACE

Les niveaux que l'on peut imaginer sont extrêmement nombreux car on dispose d'une assez grande liberté dans leur définition. Aussi les applications précises que l'on poursuit interviennent avec grand poids dans un tel choix.

Prenons à titre d'exemple, le modèle qui a été réalisé au C.E.T.A. pour l'application « traduction automatique » et qui pourrait être utilisé en vue d'autres applications.

Une description détaillée de ce modèle et des résultats auxquels il conduit se trouvent dans les références [1], [2], [3], [4].

Pour donner un aperçu très sommaire du niveau obtenu on peut dire qu'une formule à ce niveau exprime le contenu d'une phrase en utilisant la lexicographie de la langue naturelle et en se débarrassant le plus possible de la syntaxe. Ainsi les relations entre les mots qui étaient de nature syntaxique au niveau précédent sont remplacées par des relations logiques.

Une formule apparaît encore sous la forme d'une arborescence où les sommets sont porteurs d'unités lexicales et les arcs porteurs de relations.

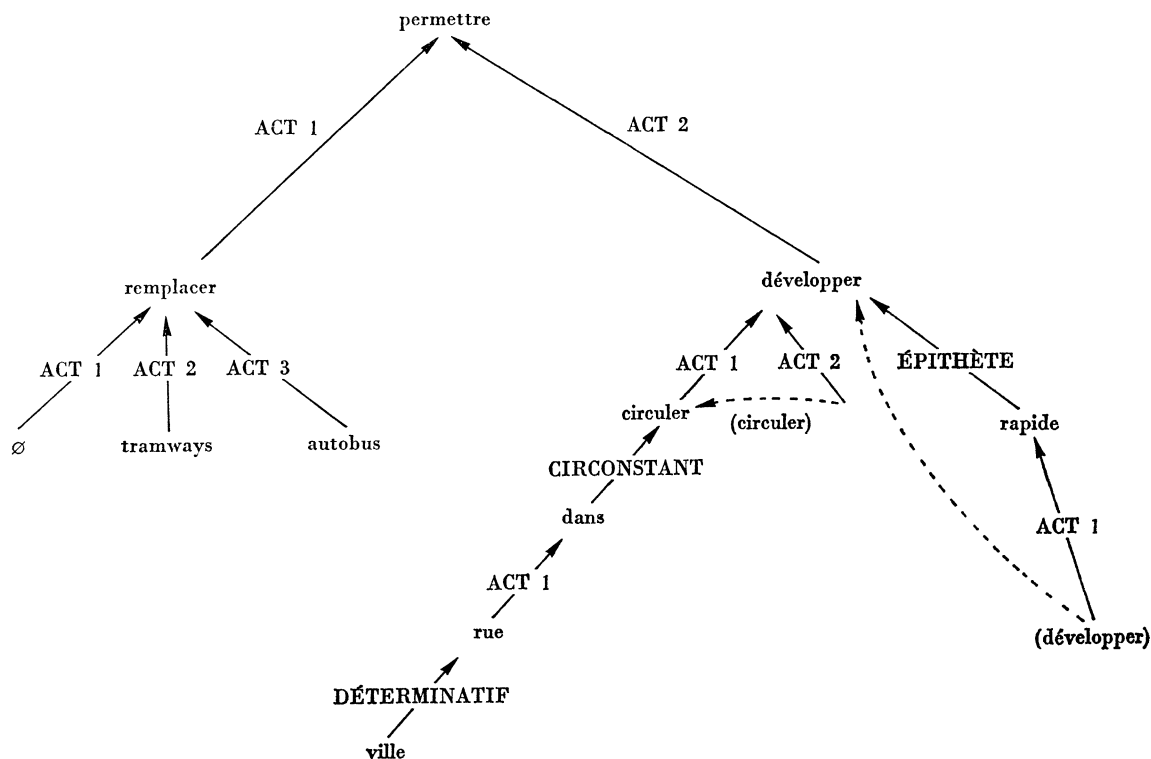
Ces relations appartiennent à l'un ou à l'autre des deux types suivants :

a) Relation indiquant la place d'un argument dans un prédicat (prédicat à une, deux ou trois places d'argument). Cette relation est notée ACT1, ACT2, ACT3, suivant la place (rappelant ainsi la notion d'actant de Tesnière).

b) Relations indiquant la dépendance qui existe entre deux prédicats. On se bornera à expliciter celles qui interviennent dans la formule associée à la phrase donnée en exemple à propos du modèle syntaxique.

Ainsi, cette phrase qui sert d'exemple :

« le remplacement des tramways par des autobus a permis un développement rapide de la circulation dans les rues de la ville » aboutit à la formule :



Comme dans le cas de la structure syntaxique, cette formule est simplifiée en ce sens que des indications de nombre, de temps (temps réel et non temps syntaxique), etc., sont aussi associées aux unités lexicales.

Toujours est-il que cette formule met en évidence les énoncés suivants :

- remplacer permet de développer,
- on remplace les tramways par des autobus,

- le développement est rapide,
  - la circulation « dans » la rue,
  - la rue « de » la ville,
- en outre, on connaît l'articulation de ces énoncés.

Il en résulte un éloignement tel de la construction syntaxique initiale que bon nombre d'autres phrases de signification équivalente et utilisant les mêmes unités lexicales aboutiraient à la même formule.

En conséquence, les phrases commençant par :

- « le fait d'avoir remplacé les tramways par des autobus... »,
  - « le fait que l'on ait remplacé les tramways par des autobus... »,
  - « que les tramways aient été remplacés par des autobus, cela a permis... »
- aussi, celles continuant par :
- « a permis de développer rapidement la circulation... »,
  - « a permis que la circulation se développe rapidement... »,
- ainsi que les voies passives admissibles.

### *Applications*

- a) La première application de ce niveau a été le système de traduction automatique du C.E.T.A. en mettant en œuvre un modèle concret d'analyse à partir du niveau syntaxique pour le russe et l'ébauche d'un modèle analogue pour le japonais et enfin un modèle concret de synthèse pour le français ;
- b) On peut de toute évidence utiliser un aller-retour de modèles d'analyse et de synthèse dans la même langue en vue de l'étude de la paraphrase ;
- c) On imagine fort bien qu'un niveau de ce genre (soit un peu plus évolué, soit un peu moins) soit le pivot (ou l'interface) de la communication homme-machine en langue naturelle. Il y a là matière à de nombreuses applications spécialisées : interrogation de fichier, documentation automatique, commande en langue naturelle, enseignement assisté par ordinateur, etc.
- d) Enfin, pourquoi pas, considérer ce niveau comme étape intermédiaire vers des buts plus ambitieux.

En quelques pages on ne peut espérer entrer dans les détails de toute l'activité concourant au traitement automatique des langues. Cette notion de niveau de langue et celle de modèle se sont révélées comme éléments moteurs de cette activité. Réunissant linguistes, mathématiciens et informaticiens autour d'un objectif commun qui ne relève pas uniquement soit des propos de la linguistique soit de l'un des autres, le traitement automatique des langues est un exemple éclatant d'activité pluridisciplinaire (ou interdisciplinaire) riche en nouvelles découvertes et en nouvelles applications.

## BIBLIOGRAPHIE

- [1] VAUQUOIS, B., VEILLON, G., "Un métalangage de grammaires transformationnelles", *Conférence Internationale sur le Traitement Automatique des Langues*, 1967 (1967 International Conference on Computational Linguistics), Grenoble, 23-25 août 1967.
- [2] VEILLON, G., "Description du langage pivot du système de traduction automatique du CETA", *T.A. Informations*, 1968, n° 1 pp. 8-17.
- [3] VAUQUOIS, B., VEILLON, G., NEDOBEJKINE, N., BOURGUIGNON, C., "Une notation des textes hors des contraintes morphologiques et syntaxiques de l'expression", *International Conference on Computational Linguistics, Stockholm*, 1-4 septembre 1969.
- [4] VEILLON, G., *Modèles et algorithmes pour la traduction automatique*, Thèse de Doctorat, Grenoble, avril 1970.