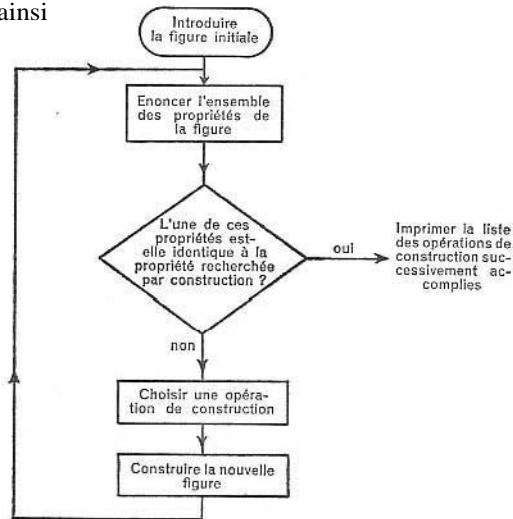


Le théorème cité se formule alors comme suit

$$\begin{aligned}
 & [\text{Diag} (D, Q)] \ \& \ (D_1 \in Q) \ \& \ (D_2 \in Q) \ \& \ (D_3 \in Q) \ \& \ (D_4 \in Q) \\
 & \ \& \ X (P_1, D_1, D_2) \ \& \ X (P_2, D_3, D_4) \\
 & \ \& \ R (D, P_1, P_2) \\
 & \ \& \ X (P_3, D, R (X (D_1, D_3), X (D_2, D_4))) \\
 & \ \& \ X (P_4, D, R (X (D_1, D_4), X (D_2, D_3))) \\
 \Rightarrow & \text{Conj Harm} (P_1, P_2, P_3, P_4).
 \end{aligned}$$

Le problème de la recherche automatique des constructions se précise alors ainsi



Ici, aux difficultés rencontrées précédemment, et qui concernent la possibilité de cyclage indéfini d'opérations, s'en ajoute une autre, celle contenue dans le rectangle « énoncer l'ensemble des propriétés de la figure ».

Ce sont les difficultés que nous avons évoquées plus haut, à l'occasion de notre discussion sur les programmes de logique automatique. Pourtant, le fait d'avoir choisi un domaine assez restreint de la géométrie élémentaire nous permet d'entrevoir une possibilité de solution. Nous y reviendrons dans les chapitres suivants.

CHAPITRE V

Le langage

Au cours des deux chapitres précédents, nous avons indiqué que l'étude de systèmes tels que les jeux ou les disciplines mathématiques pouvait être considérée comme une première approximation de l'étude du langage, dans la mesure où ils forment eux-mêmes des langages relativement simples et en tout cas complètement formalisables. Nous avons cependant noté que, en particulier dans le cas des mathématiques, la formalisation complète est rarement effectuée et pose même parfois des problèmes insolubles.

Cela n'est pas étonnant puisque, revenant pour un instant au point de vue génétique, il est certain que les aptitudes au jeu, au raisonnement, etc., se développent de pair avec l'aptitude au langage et forment avec elle un ensemble difficilement séparable en parties.

A tous points de vue, le langage est ainsi le support essentiel du développement de l'intelligence et même de son existence. Il est donc obligé que le « front linguistique » de l'intelligence artificielle joue un rôle central dans notre propos. Succès et échecs de la linguistique automatique sont typiques du pouvoir et des limitations des procédés dont dispose à ce jour la discipline dont nous nous occupons ici.

Pour simplifier les choses, nous ne considérerons que les problèmes relatifs au traitement automatique de la langue écrite et l'exemple qui soutiendra nos diverses analyses sera emprunté à la seule langue française ; pourtant, on pourra se convaincre, chemin faisant, que

ces restrictions n'affectent pas sérieusement la validité des conclusions que nous pourrions tirer.

Le fait de nous restreindre au langage écrit simplifie considérablement le problème de *l'introduction des données*. Ce qui correspondra en effet, dans les programmes linguistiques, aux « données » des programmes décrits précédemment, ce sont les textes écrits (et transformés, pour la commodité de l'automate en cartes perforées, bandes perforées, bandes magnétiques, etc., grâce à l'un des codages décrits au début du chapitre II). Ce qui correspondra à ce que nous avons appelé plus haut la « documentation de base » ce sera l'ensemble des règles de grammaire, règles sémantiques, etc.

Toute la différence avec les problèmes traités précédemment réside dans l'absence jusqu'à ce jour d'une formulation exhaustive et même seulement cohérente, de ces différents systèmes de règles.

Il saute aux yeux qu'un texte écrit présente les propriétés qui nous ont servi à définir un monoïde au chapitre II. Prenons garde cependant que cette structure de monoïde existe simultanément à plusieurs niveaux avec les alphabets suivants (on pouvait en envisager davantage en découpant plus finement)

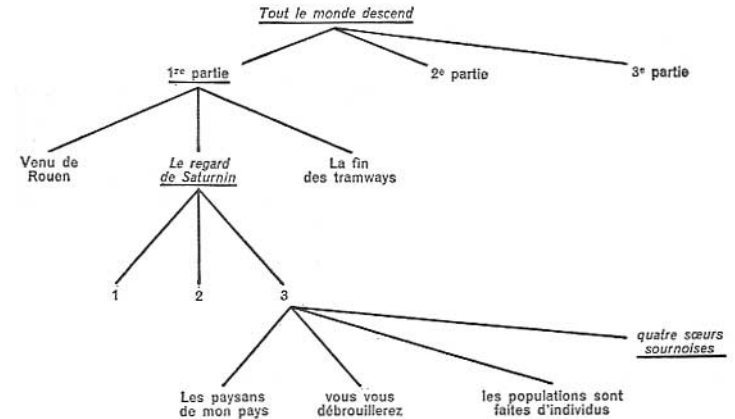
- un alphabet ordinaire ; les mots de ce monoïde sont les mots ordinaires ;
- un vocabulaire de mots ; les mots de ce monoïde sont les phrases du texte ;
- une liste de phrases ; les mots de ce monoïde sont les paragraphes ;
- une collection de paragraphes ; les mots de ce monoïde sont les chapitres.

Un même groupement de mots va donc pouvoir être considéré comme un élément d'un alphabet, ou un mot, ou une suite, suivant les niveaux auxquels on se place. On touche donc du doigt dès à présent - et sur un problème relativement élémentaire - la complexité du problème linguistique.

Cette complexité, ainsi que les moyens pour la réduire, apparaîtront mieux sur des exemples. Il nous a semblé intéressant de tirer nos exemples d'un texte unique, afin de faire mieux voir l'interprétation des problèmes et des méthodes.

UN TEXTE EXEMPLAIRE

Ce texte exemplaire sera un fragment de l'autobiographie poétique de Jean Queval : *Tout le monde descend*. Ce fragment se situe ainsi dans l'ouvrage



Il s'agit donc du dernier paragraphe de la troisième section du deuxième chapitre de la première partie de l'ouvrage ; nous le reproduisons ci-dessous *in extenso* (en y ajoutant une numérotation des phrases)

« 1. Quatre sœurs sournoises sur qui les garçons se retourneraient, s'ils l'osaient seulement, imitent la houle de la mer avec des effets de hanche. 2. Leur frère pêche à Terre-Neuve. 3. L'une d'elles a été vue avec trois garçons, à mi-falaise, entre des haies, tenant en rond quelque concile, assise à même les grosses et petites pierres dans une allée presque perdue du souvenir, tenant un concile en rond avec trois garçons, un soir de l'été. 4. A ces filles, parle ou bien ne parle pas, pour des raisons ignorées, la petite bonne de l'hôtel de la plage qui, deux fois la semaine, grossit au télescope le navire qui glisse sur la ligne d'horizon. 5. A bord, son fiancé est peut-être. 6. A tous ces êtres, Saturnin est le baromètre de l'humeur commune, comme un

jardinier de Giraudoux. 7. Cet ancien marin bricole dans les bistrots, pour une croûte et quelques verres. 8. Un mot de travers de l'un ou de l'autre, et on le chasse au balai, Saturnin. 9. Alors, au lieu d'aller dormir chez sa sueur, il découche pendant des jours. 10. C'est pendant une de ces périodes qu'eut lieu le grave concile d'une fille et de trois garçons qui probablement portait sur l'étrangeté des voies de la nature. ii. A ces moments-là, je le dis pour n'y plus revenir, tout est possible, vu le dérangement des astres. 12. Tout aussi bien, Saturnin revenu, l'ordre social se rétablit, comme par le retour de mouvement d'une grande horloge inconnue. 13. Ce n'est peut-être pas un fait de nature à simplifier la métaphysique, mais il en faut tenir compte, ou plutôt s'y soumettre, comme au temps même. 14. Le garde-champêtre le sait, Pierrot le facteur le sait, comme le savent la mère de Monsieur le Curé, et Monsieur le Curé lui-même, qui n'est pas parvenu à faire entrer cette donnée locale dans la véritable orthodoxie, au cours des promenades qu'il accomplit au bord de la mer, tout seul soir après soir. 15. La dernière fois que j'ai vu Saturnin, il pleuvait, et un couple s'était réfugié sous une porte - à bien regarder, on eût reconnu la championne de tennis et son admirateur ignoré -, et ce navigateur de la cloche, Saturnin, heureux seulement d'être là, allait son chemin, sous la pluie. 16. Voyant ce couple qui ne voulait pas être mouillé, il lui fit un discours de biais, c'est-à-dire à la troisième personne, celle qui concerne les tiers, et par là permet la généralisation. 17. « Z'ont point d'chance les estivants, c't'année », dit-il. i8. Puis ce furent les grandes tempêtes de l'automne. i9. Les dernières gens des vacances repartirent vers l'intérieur. 20. Les amoureux de six heures et demi du matin s'aiment maintenant dans le demi-jour gris venu de la mer, et le village se referme comme une huître des profondeurs, Saturnin régner. »

Ce texte représente le paragraphe entier. Il comprend (si l'on limite les phrases aux textes compris entre deux points), 20 phrases. Les phrases contiennent environ 484 mots (« environ » fait allusion à différents points de vue possibles concernant les mots tels que : « mifalaise », « c'est-à-dire », etc.). A ces mots s'ajoutent 48 virgules, 1 paire de tirets et une paire de guillemets.

Il s'agit donc d'un texte relativement court, et de composition tout à fait raisonnable. Et pourtant ce texte comporte, pour l'analyste,

un très grand nombre de difficultés - et c'est à ce titre que nous l'avons sélectionné. Nous sommes en effet bien loin du compte rendu, du procès-verbal objectif, voire du récit d'événements.

Tant que l'on demeurera au niveau du lexique ou même de la syntaxe, on ne rencontrera pas de difficultés plus grandes que pour l'analyse de tout autre texte, mais dès qu'on abordera le niveau sémantique et qu'on voudra pousser l'analyse en profondeur, on rencontrera des difficultés considérables.

Notre but n'étant pas seulement de décrire la situation actuelle dans ce qu'elle a de satisfaisant, mais aussi de souligner l'existence et l'importance des difficultés, nous développerons plus particulièrement ces derniers aspects, en utilisant toujours la représentation commode que nous offrent les automates et les programmes dont on peut les munir.

Nous allons donc décrire le traitement possible du texte par un automate, en énumérant un certain nombre de programmes, en respectant un certain enchaînement logique, mais sans insister sur les conditions rigoureuses (format d'entrée et de sortie) qu'exigerait l'enchaînement effectif dans un automate réel.

Rappelons que ces programmes mettent en oeuvre

- un algorithme général (tel l'algorithme d'analyse syntaxique) ;
- une documentation de base (telle que le dictionnaire) ;
- des données particulières (le texte à traiter) ;
- un résultat.

Nous avons attribué à chacun d'eux un symbole mnémorique comme pour les chapitres précédents.

ANALYSE LEXICALE

Cette analyse, quoique triviale est évidemment fondamentale. D'ailleurs la lexicographie se tourne désormais de plus en plus vers l'utilisation des machines qui lui sont indispensables lorsqu'elle se propose de manipuler des masses vraiment significatives de textes (i).

IL	algorithme : comparaison ;
Identification	documentation de base : lexique ;
lexicale	données : les mots.

(1) On se reportera notamment avec fruit aux travaux de j. Quemada à Besançon.

Il s'agit simplement de vérifier que les mots utilisés existent (dans le lexique de base utilisé). Par exemple, si ce lexique est le *Petit Larousse illustré*, les mots « z' » et « ânée » seront indiqués comme aberrants. Nous n'indiquerons pas ici comment ces anomalies pourraient être prises en charge par un programme spécialisé (dont la documentation de base serait de nature phonétique).

CL algorithme : comptage ;
Comptage documentation de base : programme A. ;
lexical données : les mots.

Il s'agit d'attribuer à chaque mot identifié un registre qui augmente d'une unité chaque fois qu'on rencontre à nouveau le même mot. L'application de $A_1 - A_2$ au texte choisi T donne la statistique suivante (chaque mot est suivi du nombre de ses apparitions dans le texte

le	16	des	6	ce	4
la	15	d'	6	ces	4
de	12	Saturnin	6	dans	4
à	11	au	5	garçon	4
l'	9	est	5	seulement	4
une	8	il	5	etc.			
un	8	les	4				
les	7							

L'examen, même rapide, de cette statistique en montre les limitations « de » et « d' », « se » et « s' », etc., sont comptés comme des mots différents. Il en va de même de : « est », « être », « furent », etc., ou de « fille » et « filles ». Par contre, on a identifié des apparitions de « la », « leur », etc., qui ne correspondent pas au même être linguistique.

Parmi les programmes de regroupement, nous pouvons définir dès maintenant

RL algorithme : élimination des élisions ;
Regroupement documentation de base : liste de règles de morphophoné-
 mique ;
lexical données : mots.

La mise en œuvre de RL sur T fera d'ailleurs sortir de la liste des mots identifiés, le « c' » de « ç't'ânée » auquel ne correspond aucune élision visible.

Mais bien entendu le programme de regroupement le plus important est le suivant

RP algorithme : identification de mots qui appartiennent à
Regroupement un même paradigme grammatical ;
paradigmatique documentation de base
 - règles des genres et des nombres ;
 - listes de pluriels irréguliers, tables de conjugaisons
 (régulières ou non) ;
données : mots.

On conviendra de remplacer les formes fléchies par la forme singulier masculin ou la forme infinitive, suivant la nature grammaticale des mots.

Après insertion de RL-RP entre IL et CL, on obtient la nouvelle statistique

le	40	ce	12
de	24	être	9
à	16	Saturnin	6
un	16	etc.		

Le document de base qui permet la mise en œuvre du programme RP peut servir également à produire des couples (mots, nature paradigmatique) tel que

imitent : 3e personne du pluriel de l'indicatif présent du verbe imiter.

Plus généralement, nous utiliserons le programme d'identification grammaticale

IG algorithme : recherche de la catégorie grammaticale et, le cas échéant,
 identification du paradigme ;
documentation de base : dictionnaire grammatical ;
données : mots ;
résultat : couples (mot, catégorie ou paradigme).

On notera que le résultat n'est pas nécessairement unique.

C'est ainsi que le quatrième mot de la douzième phrase : « revenu » introduit dans le programme IG produira deux résultats différents (revenu, substantif masculin singulier) et (revenu, verbe revenir participe passé masculin singulier).

Lorsqu'un homme - ou même un enfant - effectue ce genre d'analyse, il élimine automatiquement les possibilités visiblement contra-

dictoires avec le reste du texte. L'automate ne le fait pas, ou plutôt il ne peut le faire qu'après la mise en oeuvre d'autres programmes spécialisés.

ANALYSE SYNTAXIQUE

Ceci veut dire que l'application du programme IG a pour effet principal de rendre manifestes les ambiguïtés du langage. Ces ambiguïtés ne peuvent pas être levées au niveau de la grammaire (au sens usuel du terme, c'est-à-dire déclinaison, conjugaison, etc.) ; il faut faire appel à la syntaxe. Pour bien marquer ce qui fait l'importance et la force de la syntaxe, considérée en tant que système de contraintes, nous utiliserons une *représentation graphique* des contraintes ainsi manifestées. Ce type de représentation est maintenant fort à la mode et on en possède plusieurs systèmes, plus ou moins équivalents.

Le choix, important pour les linguistes, n'est pas significatif en ce qui concerne notre présent effort, qui est seulement un effort d'illustration des méthodes. Aussi, le choix que nous proposons n'est-il nullement impératif. Nous utilisons la méthode due à L. Tesnière qui a, en particulier pour nos lecteurs, le mérite d'être exposé dans un ouvrage rédigé en français (1).

Il s'agit de marquer la contrainte syntaxique qui lie les éléments d'un syntagme en liant physiquement les mots par un trait. Le graphe ainsi obtenu s'appelle « stemma ». Les stemmas dénotent une sorte de hiérarchie engendrée par une hiérarchie des catégories grammaticales, hiérarchie que l'on définit par les inégalités

|verbe| > |substantif| > |adjectif| > |adverbe|

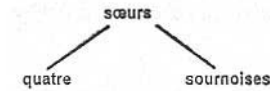
Des règles complémentaires permettent de relier entre elles des propositions élémentaires dont l'articulation est du type : proposition relative, subordonnée, etc.

Dans un tel système syntaxique, les lois ont l'apparence des formules de la chimie organique

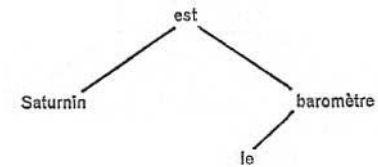


(1) L. TESNIÈRE, *Éléments de syntaxe structurale*, Klincksieck, 1959. De nombreux auteurs utilisent maintenant la notation de Tesnière, sans toujours le mentionner !

C'est ainsi qu'au début de la première phrase la suite « quatre sœurs sournoises » donne



Phrase 6, après la virgule, la suite « Saturnin est le baromètre » donne



Ces formules pseudo-chimiques n'expriment rien d'autre, bien entendu, que les règles classiques de la grammaire, puisqu'elles visualisent la construction progressive de groupes nominaux, de groupes verbaux, etc. Elles permettent, en particulier de prendre aisément en charge les règles d'accord en genre et en nombre, l'articulation des propositions de différents types, etc.

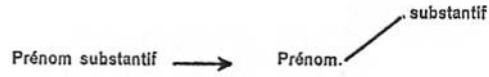
Nous ajouterons donc à la liste de nos programmes

CS algorithme : construction des stemmas relatifs à un groupement de mots donnés ;
documentation de base : syntaxe ;
données : groupes de couples (mot, catégorie grammaticale) ;
résultat : stemmas.

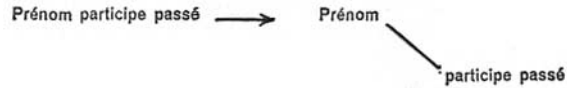
Succès ou échecs de la construction aboutissent à conserver ou à éliminer certains des couples (mot, catégorie) ; on peut donc construire le programme d'élimination des ambiguïtés

EA algorithme : élimination de couples (mot, catégorie) ;
documentation de base : IG CS ;
données : liste de couples ;
résultat : liste restreinte de couples.

C'est ainsi que le fragment « Saturnin revenu », tiré de la phrase 12, étant compris entre deux virgules, doit être connexe. Or il n'existe aucune règle de grammaire permettant d'écrire



par contre il existe une règle qui donne

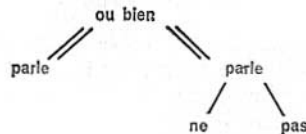


On élimine ainsi l'ambiguïté relative au mot « revenu ». On imagine sans peine qu'une syntaxe efficace doit comporter un très grand nombre de telles règles dont certaines très rarement usitées mais tout de même indispensables à la cohérence de la langue en tant que système. L'ouvrage de L. Tesnière ne donne que quelques-unes de ces règles et, à vrai dire, on ne dispose, en ce moment, pour aucune langue, d'un ensemble de règles suffisamment proche de l'exhaustivité. Cela veut dire que dans certains cas l'analyse syntaxique ne pourra être menée jusqu'au bout ou que, au contraire, plusieurs constructions seront proposées pour un même fragment.

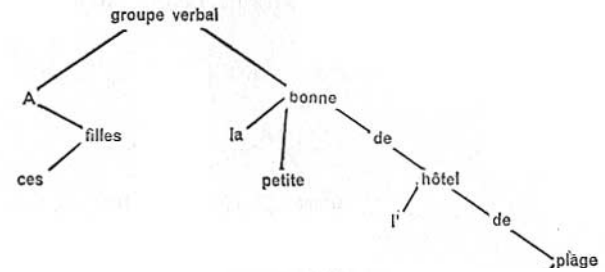
Pour l'exemple que nous avons choisi, aucune de ces difficultés n'apparaît et chaque phrase donne naissance à un ou plusieurs stemmas. Nous donnons ci-dessous quelques exemples (la numérotation est celle des phrases (de 1 à 20)

Examinons tout d'abord la *phrase 4* qui présente plusieurs particularités intéressantes

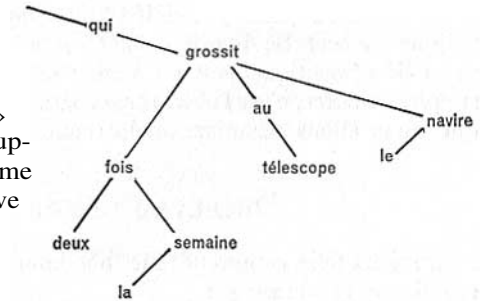
Tout d'abord le « sommet » du graphe n'est pas un mot, mais un groupe de mots



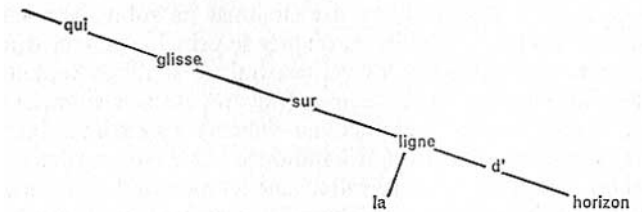
A ce groupe sont rattachés un graphe sujet et un graphe complément, mais avec une inversion



le substantif « bonne » reçoit un qualificatif supplémentaire sous la forme d'une proposition relative

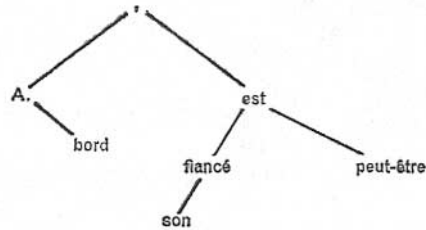


le mot « navire » à son tour est qualifié à l'aide d'une proposition relative



d'où finalement une structure globale extrêmement télescopique.

Un autre exemple intéressant est celui de la *phrase 5* où l'expression « à bord », qui qualifie le verbe « est », ne peut être acceptée *avant* le verbe que grâce à la présence de la virgule, ce qui donne le graphe



Mais cette phrase présente encore un autre intérêt : c'est la présence de l'adjectif possessif « son » que notre construction syntaxique permet de lier à « fiancé » mais qui se trouve évidemment lié sémantiquement à « bonne ». Établir de telles connexions supplémentaires, c'est l'objet des programmes d'analyse sémantique, que nous allons examiner maintenant.

ANALYSE SEMANTIQUE

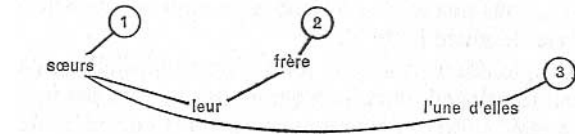
Pour mieux faire comprendre le fonctionnement de ces programmes, examinons la phrase 4

« A ces filles parle ou bien ne parle pas, pour des raisons ignorées, la petite bonne de l'hôtel de la plage qui, deux fois la semaine, grossit au télescope le navire qui glisse sur la ligne d'horizon. »

De toute évidence, la structure des stemmas ne suffit pas à épuiser la cohérence du texte. D'ailleurs, d'après le principe de construction, les sommets sont toujours des verbes, qui ne sont pas toujours les mots les plus significatifs en stemma (notamment s'ils sont des auxiliaires). Une première retouche au schéma syntaxique est alors apportée par la construction (déjà indiquée par Tesnière) des liaisons « anaphoriques », c'est-à-dire rattachant les pronoms personnels ou possessifs aux individus ou objets qu'ils représentent. Il s'agit donc d'un nouveau programme :

CA algorithme : tracé de nouvelles connexions entre mots ;
 documentation de base : loi des connexions anaphoriques ;
 données : phrases du texte ;
 résultat : liens horizontaux entre stemmas.

Appliqué au texte T, CA donne, par exemple, le résultat suivant



La simple application des connexions anaphoriques ne suffit pas à expliciter toutes les connexions du même type. C'est ainsi que « ces », de « ces filles », et aussi de « ces êtres », réclame une connexion que les règles élémentaires ne donnent pas.

Il faut donc mettre en oeuvre un nouveau programme qui va plus profondément dans l'analyse sémantique, par exemple

CH algorithme : mise à jour de connexions hiérarchiques ;
 documentation de base : classifications diverses ;
 données : mots ;
 résultats : nouveaux liens anaphoriques.

Dans les classifications en question, on voit notamment que les soeurs sont nécessairement des filles, que les filles, bonne et fiancé, sont des êtres, etc.

Remarquons que, sauf les exceptions des classifications des espèces vivantes, ou des minéraux, des composés chimiques, etc., de telles classifications ne sont pas, pour l'instant, complètes.

Enfin, de nouvelles connexions plus générales, de type sémantique, peuvent apparaître, telle celle qui, dans les phrases 4 et 5, relie « navire » et « à bord ». Elles seront mises en évidence à l'aide d'un programme dont les documents de base pourraient être des ouvrages du type « recueil de mots français par catégories », dictionnaires analogiques, etc., d'où le programme qui suit.

CS algorithme : établissement de connexions sémantiques ;
 documentation de base : liste de mots classés par sujet ;
 données : mots ;
 résultat : nouvelles connexions entre stemmas.

De nouveaux programmes sémantiques plus complexes ou plus abstraits doivent être également mis en oeuvre. Pour ne pas surcharger notre analyse, nous ne citerons ici que ceux dont la nécessité s'impose pour l'analyse de notre texte T.

Remarquons que désormais nous nous situons au-delà de ce qui est effectivement développé. De tels programmes ne sont pas impossibles à construire mais coûtent évidemment un effort considérable qui n'a été qu'abordé dans les différentes équipes qui travaillent aujourd'hui dans les divers domaines de la linguistique automatique.

ST algorithme : établissement de connexions spatiales ou temporelles ;
 documentation de base : règles de grammaire sémantique spéciales (relatives notamment aux aspects du verbe) ;
 éléments : mots ou syntagmes ;
 résultat : nouvelles connexions entre polystemmas.

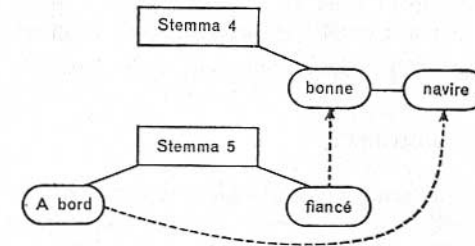
Cet algorithme met en évidence des relations de voisinage, de succession, etc. Il doit être distingué de CH en raison de l'utilisation de règles particulières

Ainsi, le « puis » de i8 assure la liaison avec le groupe 16-17. Les passés simples « furent » et « repartirent » unissent i8 et i9. Par contre, le présent « aiment » de 2o affirme la séparation temporelle de 2o d'avec les groupes précédents.

Tout ceci nous permet donc d'achever d'établir la connexion globale de l'ensemble du texte, ainsi d'ailleurs que le manifestera plus clairement encore le programme suivant

CG algorithme : substitution aux deux stemmas d'un symbole (par ex. le numéro du stemma) auquel se rattachent les mots reliés anaphoriquement, puis fabrication d'un graphe où ces liaisons, ainsi que les autres liaisons sémantiques qui viennent d'être définies sont explicitées ;
 documentation de base : l'ensemble des programmes précédents ;
 données : texte complet ;
 résultat : graphe récapitulatif des connexions textuelles.

Appliqué à T, CG donne le résultat suivant



Dans de nombreux cas, l'établissement du métastemma permettra de résoudre les tâches pour lesquelles un automate avait été nécessaire.

Pourtant, lorsqu'on veut aller plus loin dans la compréhension, une deuxième phase de l'analyse peut être nécessaire. Cette phase que nous allons décrire maintenant, en utilisant toujours notre texte T comme illustration, nous fera passer du métastemma au diagramme, en passant par l'intermédiaire d'une formalisation.

Ici encore, les diverses étapes seront présentées en termes de programmes exécutables par un automate. Nous utiliserons successivement

II algorithme : comparaison ;
 Identification des individus documentation de base : listes d'individus ;
 données : noms propres, noms de personnes, etc.
 résultat : couples individus-symboles.

II (T) -* quatre so.urs, un fiancé, trois garçons, etc.

ID algorithme : comparaison ;
 Identification des décors documentation de base : listes des décors, lieux, etc. ;
 données : syntagmes ;
 résultat : listes de couples (figure, symbole).

ID (T) -* bord de la mer, falaise bistrot, etc.

IO algorithmme : comparaison ;
 Identification des objets documentation de base : listes d'objets ;
 données : mots ou groupes de mots ;
 résultat : couples (groupes de mots, symboles).
 IO (T) -> pierre, télescope, navire, etc.

et de même on construira

un programme IT pour l'identification des instants

—	IE	—	événements
—	IP	—	prédicats élémentaires
—	IR	—	relations

puis un programme CP utilisant les résultats des programmes précédents, ainsi que les programmes construits dans le cadre de la logique des prédicats (au sens du chapitre précédent), pour construire des prédicats complexes.

L'ensemble de ces programmes d'identification appliqué à T donne le résultat suivant

a) Individus

I ₁ : sœurs	I ₉ : jardinier	I ₁₇ : mère
I ₂ : garçons	I ₁₀ : Giraudoux	I ₁₈ : curé
I ₃ : frère	I ₁₁ : marin	I ₁₉ : couple
I ₄ : garçons	I ₁₂ : sueur	I ₂₀ : championne
I ₅ : filles	I ₁₃ : fille	I ₂₁ : admirateur
I ₆ : bonne	I ₁₄ : garçons	I ₂₂ : estivants
I ₇ : fiancé	I ₁₅ : garde-champ.	I ₂₃ : gens
I ₈ : Saturnin	I ₁₆ : Pierrot	I ₂₄ : amoureux
		I ₂₅ : huître

b) Prédicats

P₁ (≡ (I₁)) = 4
 P₂ (I₁) = sournoises

P₃ sera défini lors de l'analyse de la phrase n° r
 P₄ (I₁) = frère
 P_s (I₃) = pêcher
 P₆ (I₁) (≡ E) = l'une d'elles - I₂₆
 P₇ (I₂₆) = être vu
 P₈ (I₂₆) = tenir (un concile)
 P₉ = être assis

c) Actions

A ₁ : se retourner	A ₁₆ : dire	A ₃₁ : vouloir
A ₂ : oser	A ₁₇ : rétablir	A ₃₂ : faire un discours
A ₃ : imiter	A ₁₈ : simplifier	A ₃₃ : concerner
A ₄ : pêcher	A ₁₉ : falloir	A ₃₄ : permettre
A ₅ : voir	A ₂₀ : tenir compte	A ₃₅ : repartir
A ₆ : tenir	A ₂₁ : soumettre	A ₃₆ : aimer
A ₇ : parler	A ₂₂ : savoir	A ₃₇ : refermer
A ₈ : grossir	A ₂₃ : parvenir	A ₃₈ : régner
A ₉ : glisser	A ₂₄ : (faire) entrer	
A ₁₀ : bricoler	A ₂₅ : accomplir	
A ₁₁ : chasser	A ₂₆ : pleuvoir	
A ₁₂ : aller	A ₂₇ : se réfugier	
A ₁₃ : dormir	A ₂₈ : regarder	
A ₁₄ : découcher	A ₂₉ : reconnaître	
A ₁₅ : porter	A ₃₀ : aller	

d) Événements :

E₁ : houle
 E₂ : effet
 E₃ : mot
 E₄ : concile
 E₅ : ordre social
 E₆ : retour

e) Lieux :

L₁ : Terre-Neuve
 L₂ : à mi-falaise
 L₃ : entre des haies
 L₄ : bistrot
 L₅ : chez sa sueur

f) Objets

O₁ : mer
 O₂ : hanche
 O₃ : baromètre
 O₄ : croûte
 O₅ : verre
 O₆ : balai
 O₇ : horloge

g) Modes :

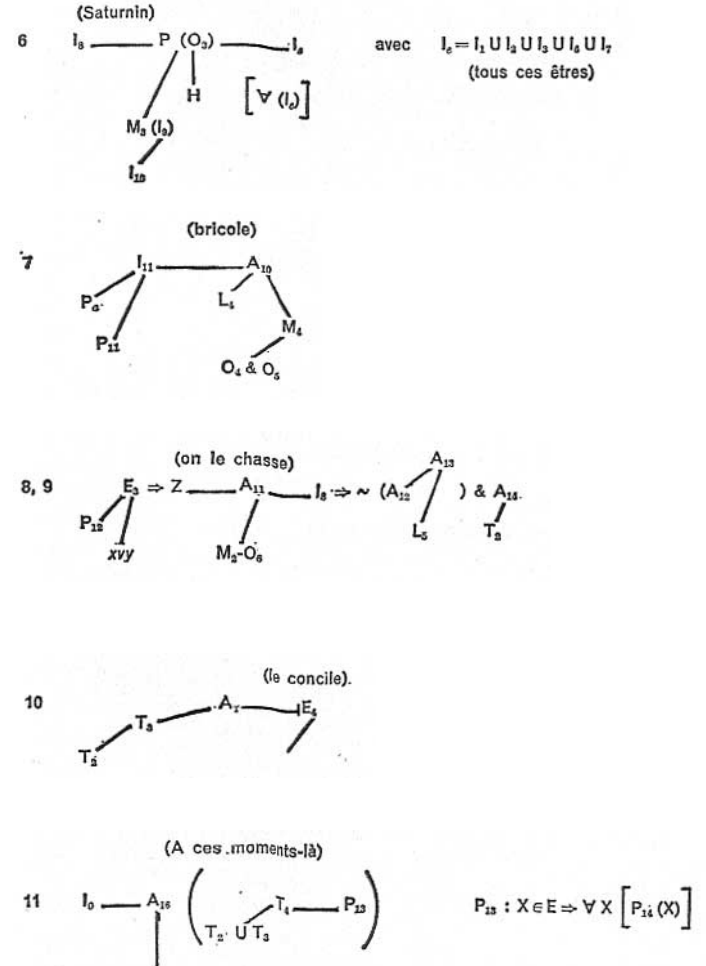
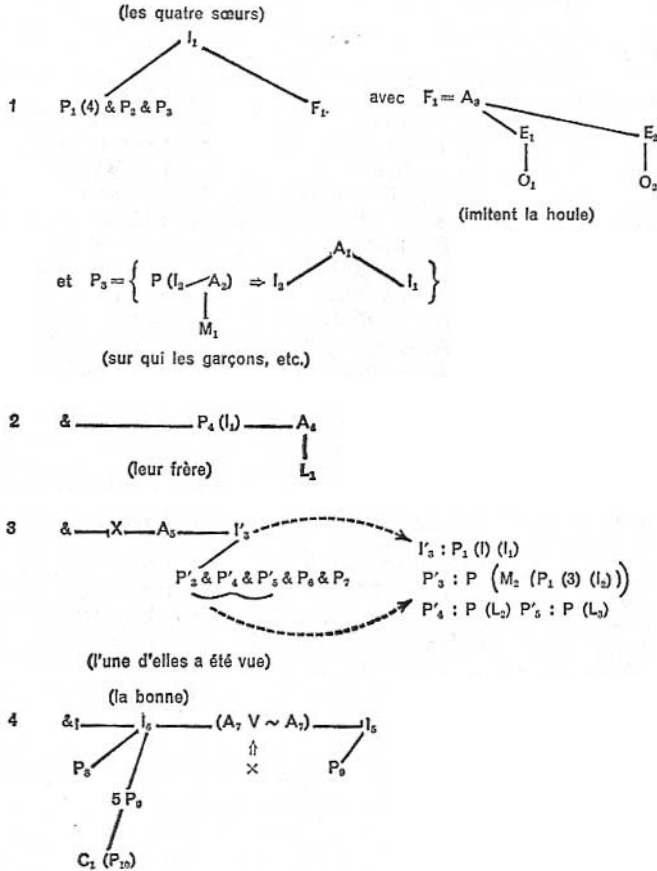
M₁ : seulement
 M₂ : avec
 M₃ : comme
 M₄ : pour

h) Temps

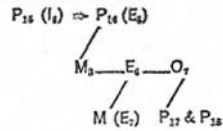
T₁ : un soir de l'été
 T₂ : pendant des jours

CONSTRUCTION DE DIAGRAMMES

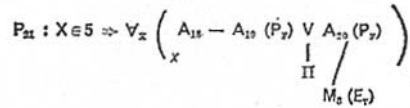
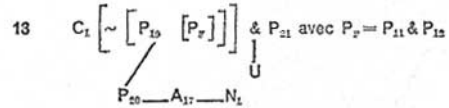
En appliquant le programme CP, on obtient alors une représentation de T sous forme d'une proposition qui se décompose en 20 prédicats reliés par la conjonction &, de la façon suivante



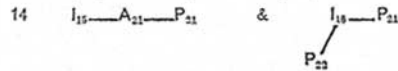
12 (Saturnin revenu)



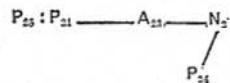
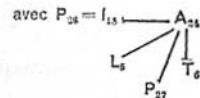
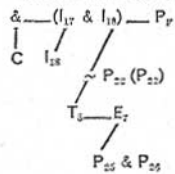
(ce n'est pas un fait de nature à ...)



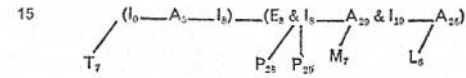
(le garde-champêtre le sait) (Pierrot le facteur)



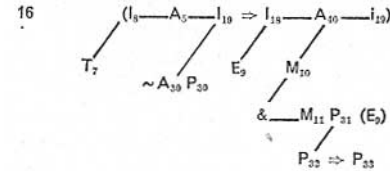
(Monsieur le curé)



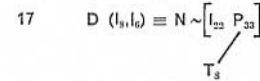
(la dernière fois que j'ai vu Saturnin)



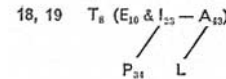
(il fit un discours de biais)



(« Z'ont point d'chance ... »)



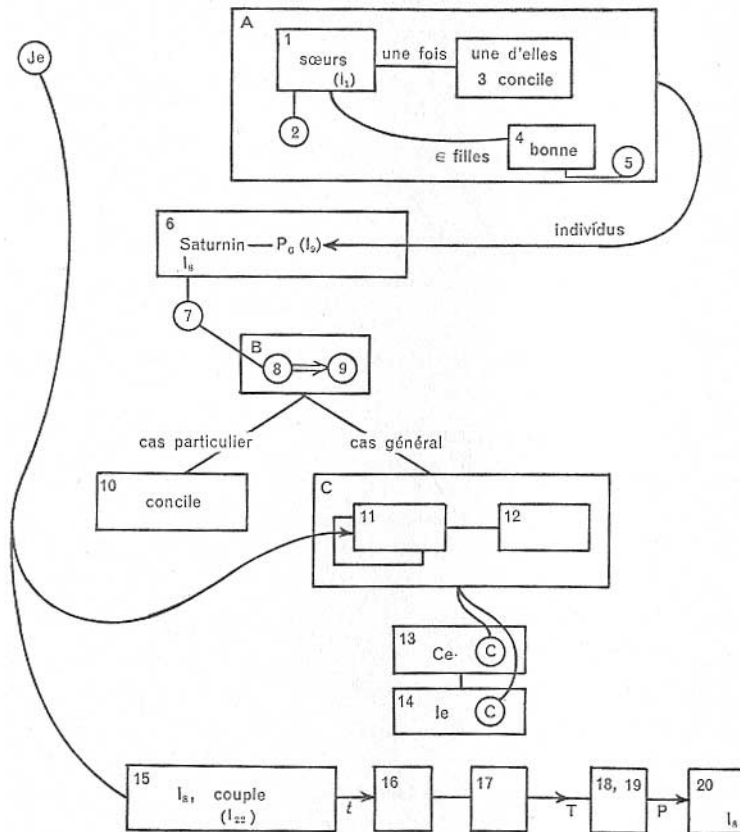
(Puis ce furent)



(les amoureux de six heures)



Une représentation plus concentrée du texte s'obtiendra alors facilement (en regroupant convenablement certains prédicats)



EXPLOITATION DES ANALYSES

Les analyses lexicales sont exploitables de bien des points de vue. Par exemple, les premiers essais de traduction automatique faisaient usage d'une telle analyse, suivie de l'utilisation de dictionnaires simples. Il s'agissait par conséquent d'un simple mot à mot (1). On utilise aussi ces analyses pour l'établissement de *concordances* qui établissent l'existence de rapprochement plus ou moins fréquents entre certains mots, enfin pour l'analyse stylistique (2).

Les analyses syntaxiques forment l'essentiel du domaine actuel de la linguistique automatique. De nombreuses équipes aux U.S.A., en U.R.S.S., en France, en Italie, en Angleterre, etc., se sont attaquées à ce problème. Dans la plupart des cas l'objectif à atteindre était la traduction automatique. Dans certains cas, cependant, l'analyse syntaxique était un but en soi (par exemple dans les études menées sous la direction de Z. Harris à l'Université de Pennsylvanie). L'expérience a montré que la mise au point de grammaires réellement complètes était un travail immense et, jusqu'à présent, inachevé. Mentionnons cependant les progrès réalisés par le Japonais Kubo à l'Université Harvard (il s'agit de l'explicitation - à l'intention d'un automate - de la grammaire anglaise ; les grammairiens prétendent souvent que la grammaire anglaise tiendrait en quelques pages ; cependant les règles mises à jour par Kubo sont au nombre de plusieurs milliers).

Le niveau sémantique a été abordé suivant des voies assez diverses. Les uns, comme S. Ceccato, à l'Université de Milan, n'ont pas voulu séparer l'aspect syntaxique et l'aspect sémantique dans l'analyse du langage. Leur intention, plus psychologique que linguistique, se manifeste bien par le nom *Adamo II* qu'ils avaient donné à leur projet d'automate.

L'attaque la plus massive sur le front sémantique est celle qu'ont menée les documentalistes. Leur but était double. Dans une première phase, ils voulaient définir un langage artificiel (artificiel par sa gram-

(1) Voir par exemple le livre d'E. Delavenney cité en bibliographie.

(2) On trouvera de nombreuses illustrations de ceci dans les ouvrages de P. Gurxaun. Voir aussi J. Cohen, *Structure du langage poétique*, Flammarion, 1966.

maire, mais naturel par son vocabulaire) qui pût décrire les documents (et aussi les questions auxquelles ces documents devraient fournir une réponse). Dans une seconde phase ils voulaient construire un programme qui donne à un automate la possibilité, à la lecture du document (ou au moins de son résumé) de construire la phrase ou les phrases qui, dans le langage artificiel ainsi défini, expriment le contenu.

Dans une première étape les langages artificiels furent totalement dépourvus de grammaire. Ils ne comportaient que des *mots clés*. Puis on introduisit des relations qui permirent la constitution de *phrases clés* (1).

Mais la vraie difficulté n'était pas dans la constitution d'une syntaxe artificielle, mais dans la tentative d'exprimer, d'une façon ou d'une autre, la signification et les rapports de signification.

De nombreux essais furent tentés dont plusieurs sont maintenant en application (2).

D'autres essais intéressants sont ceux qui se sont efforcés de manier des fragments très spécialisés du langage naturel, de façon à réduire autant que possible l'étendue du champ sémantique correspondant. On en a vu un exemple dans l'article de Slagle que nous avons cité, à titre d'exemple, dans notre introduction.

Dans tous les cas les travaux en cours sont loin d'être suffisants pour qu'on dispose dès à présent d'un réseau de relations sémantiques vraiment efficace. On peut cependant penser que si les recherches se poursuivent la situation ira en s'améliorant régulièrement.

(1) P. BRAFFORT et A. LETTOV, Des mots clés aux phrases clés, *Bulletin des bibliothèques de France*, 1959, P. 383.

(2) Voir l'article de C. WASTON, Information retrieval, dans *Advances in Computers*, 5, 1965.

CHAPITRE VI

La complexité

Simulation du joueur, du mathématicien (dans certaines de ses activités les plus simples), du sujet parlant (avec également des restrictions), au cours des trois chapitres précédents, nous avons montré comment les automates pouvaient prendre en charge la résolution de problèmes de plus en plus difficiles et que l'on situe ordinairement dans le domaine des activités intellectuelles.

Parvenus à ce point, avons-nous le droit de dire que nous avons effectivement démontré l'existence et l'efficacité d'une « intelligence artificielle » ? Pas exactement : car nous avons pris le soin, pour tous les cas traités, de donner une analyse assez détaillée des programmes que les automates devraient exécuter ; mais, dans la plupart des cas, nous n'avons pas discuté de ce que représentaient quantitativement ces programmes du point de vue de la capacité et de la vitesse dont ces automates devraient être pourvus. Dans plusieurs cas il s'agissait d'ailleurs de schémas de programmes, et non pas de programmes effectivement écrits.

Cela veut dire, certes, que l'intelligence artificielle est possible *en principe*. Et puisque, à plusieurs reprises, nous avons fait allusion à des expériences déjà effectuées (jeu de Go-Bang, démonstrations de propositions logiques, traduction du russe vers l'anglais), nous savons même que ce principe est devenu, en diverses occasions, un *fait*. Mais à aucun moment nous n'avons touché du doigt les *limites* de ce qui est effectivement possible (de ce qui est possible aujourd'hui comme de ce qu'il est raisonnable de prévoir pour demain).