

université de nancy I
u.e.r. de mathématiques

Sc. N. 75 / 74 A

contribution
à la compréhension automatique
du discours continu

mise en oeuvre d'un système paramétrable
aux niveaux morphologique, syntaxique
et sémantique

thèse

présentée pour l'obtention
du doctorat de spécialité
mathématiques appliquées
(informatique)



le 21 novembre 1975

par

jean-marie pierrel

jury :

BIBLIOTHEQUE SCIENCES NANCY 1



D 0952031216

président

c. pair

examineurs

j.c. derniamé

j.y. gresser

j.p. haton

g. pérennou

invité

m. griffiths

université de nancy I
u.e.r. de mathématiques

Sc. N. 75 / 74^A

contribution
à la compréhension automatique
du discours continu

mise en oeuvre d'un système paramétrable
aux niveaux morphologique, syntaxique
et sémantique

thèse

présentée pour l'obtention
du doctorat de spécialité
mathématiques appliquées
(informatique)



le 21 novembre 1975

par

jean-marie pierrrel

BIBLIOTHEQUE SCIENCES NANCY 1



D 0952031216

jury :

- | | |
|------------|---------------|
| président | c. pair |
| examineurs | j.c. derniamé |
| | j.y. gresser |
| | j.p. haton |
| | g. pérennou |
| invité | m. griffiths |

université de nancy I
u.e.r. de mathématiques

contribution
à la compréhension automatique
du discours continu

mise en oeuvre d'un système paramétrable
aux niveaux morphologique, syntaxique
et sémantique

thèse

présentée pour l'obtention
du doctorat de spécialité
mathématiques appliquées
(informatique)

le 21 novembre 1975

par

jean_marie pierrel

jury :

président	c. pair
examineurs	j.c. denniame
	j.y. gresser
	j.p. haton
	g. pérennou
invité	m. griffiths

Que Monsieur le Professeur C. PAIR, Directeur de l'Institut Universitaire de Calcul Automatique de Nancy, veuille trouver ici l'expression de ma profonde reconnaissance pour m'avoir proposé ce travail et pour l'intérêt manifesté tout au long de sa réalisation. Il me fait l'honneur de présider ce Jury et je tiens aujourd'hui à lui exprimer ma profonde gratitude pour la clarté de son enseignement et la formation qu'il m'a donnée.

Monsieur le Professeur G. PERENNOU, Directeur du C.I.C.T. à Toulouse, m'a fait l'honneur de s'intéresser à ce travail. Je l'en remercie vivement ainsi que d'avoir accepté de participer à ce Jury.

Je remercie Monsieur J.V. GRESSER, Ingénieur des Télécommunications, qui a bien voulu accepter de prendre connaissance de mon travail et de siéger à ce Jury.

Tous mes remerciements vont également à Monsieur J.P. HATON, Maître de Conférences à l'Université de Nancy I, pour les conseils et le soutien qu'il n'a cessé de me prodiguer depuis le début de cette étude par de nombreuses et fructueuses discussions. Sans le travail considérable qu'il a effectué depuis plusieurs années sur la reconnaissance vocale, cette étude ne se serait sans doute pas faite à Nancy.

J'adresse aussi tous mes remerciements à Monsieur J.C. DERNIAME, Maître de Conférences à l'Université de Nancy I, pour la formation qu'il m'a donnée et pour sa présence dans ce Jury.

Je remercie également Monsieur le Professeur M. GRIFFITHS du Département Informatique de l'I.U.T. de Nancy qui a bien voulu accepter notre invitation à ce Jury.

J'exprime enfin ma sincère gratitude à toutes les personnes qui ont contribué, par leur gentillesse et leur compétence, à la réalisation de cet ouvrage : en particulier Mademoiselle FRANC pour la frappe du manuscrit et toute l'équipe du secrétariat de l'I.U.T. Informatique pour le tirage et la mise en page.

à mes parents

à Elisabeth

et à notre fille Marie

S O M M A I R E

Chapitre III : Utilité des Entrées vocales en machine

3.1. Intérêt de la parole par rapport aux autres modes de communication	24
3.2. Différents types d'applications	25

PARTIE B : Un système de reconnaissance du discours continu basé sur
un analyseur syntaxiqueChapitre I : Présentation générale du système

1.1. Place de cette réalisation dans un système complet de reconnaissance	29
1.2. Position du problème	30
1.3. Le système acoustique	31
1.3.1. Principe général	31
1.3.2. Résultats fournis par le système acoustique	33
1.4. Structure générale du système	34
1.4.1. Idées directrices	34
1.4.2. Schéma général du système	36

Chapitre II : Données et résultats du système

2.1. Représentation de la phase prononcée	39
2.1.1. Informations acoustiques	39
2.1.2. Représentation de la phrase	40
2.2. Définition du langage	43
2.2.1. niveau syntaxique : la grammaire	43
2.2.2. niveau morphologique : le dictionnaire	45
2.2.3. niveau sémantique	49
2.3. Définition des résultats souhaités	50
2.3.1. Différence entre reconnaissance et compréhension	50
2.3.2. Le système de reconnaissance considéré comme un processus provoquant une action.	51

Chapitre III : L'analyseur syntaxique

3.1. Rappels au sujet de l'analyse syntaxique	53
3.1.1. Rappels de notions utilisées dans la suite	53
3.1.2. Définition de l'analyse syntaxique	55
3.1.3. Analyse syntaxique descendante	55
3.1.4. Analyse syntaxique ascendante	56
3.1.5. Solution retenue	58
3.2. Principales différences par rapport aux problèmes de compilation	59
3.2.1. Définition des terminaux	59
3.2.2. Inadaptation des algorithmes permettant de supprimer l'indéterminisme	60
3.2.3. Possibilités d'obtenir plusieurs résultats distincts	61
3.3. Module d'analyse syntaxique réalisé	62
3.3.1. Réduction de l'indéterminisme	62
a) par l'analyseur lexicographique	
b) par calcul de probabilité ou score de réussite	
3.3.2. Suppression de l'indéterminisme et de la récursivité	66
3.4. Description fonctionnelle de l'analyseur syntaxique	68
3.4.1. Création de la motrice IN de la relation	68
3.4.2. Procédure d'analyse syntaxique	69
3.4.3. Procédure d'émission des hypothèses	71
3.4.4. Procédures de services utilisées	72

Chapitre IV : L'analyseur lexicographique ou morphologique

4.1. But de l'analyseur lexicographique	74
4.1.1. Données en rentrée	74
4.1.2. Résultat fourni	75
4.2. Généralisation de l'analyseur : liberté par rapport à la grammaire	76
4.2.1. exposé du problème dans le cas de la reconnaissance de la parole	76

4.2.2. solution adoptée : analyse syntaxique pas à pas	78
4.3. Description fonctionnelle de l'analyseur lexicographique	80

Chapitre V : Le module de reconnaissance de mots

5.1. Introduction	84
5.2. Principales difficultés	84
5.3. Les méthodes de recherche lexicale	88
5.4. Présentation de la solution mise en oeuvre	91
5.5. Description fonctionnelle du module de reconnaissance de mots	95
5.6. Exemples de reconnaissance	98
5.6.1. cas de terminaux vrais	98
5.6.2. cas de terminaux de type lexique	100

Chapitre VI : Résultats fournis par le système

6.1. Phrase bien reconnue et différents types d'ambiguïtés	103
6.2. Nécessité d'une procédure de dialogue propre à chaque application	106

PARTIE C : Application à un langage donné

Chapitre I : Définition du langage

1.1. Choix du langage : langage de communication avec un standard téléphonique	109
1.2. Définition syntaxique	111
1.2.1. la grammaire	111
1.2.2. le lexique	111
1.3. Informations sémantiques	114
1.4. Validité du langage	115

Chapitre II : La procédure de dialogue propre à l'application

2.1. But à atteindre	116
2.2. Les différents cas à considérer	117
2.2.1. Phrase bien reconnue	117
2.2.2. Phrase reconnue avec ambiguïté	117
2.2.3. Phrase non reconnue : récupération des erreurs	118
2.3. Réalisation de la procédure de dialogue	118
2.3.1. Traitement des réponses en cours de dialogue	118
2.3.2. Exemple de dialogue possible	119

Chapitre III : Résultats obtenus en temps différé

3.1. Exemples de phrases reconnues	122
3.2. Exemples de résultats du système complet	122
3.2.1. Présentation et discussion de résultats du système complet	122
3.2.2. Exemples de résultats intermédiaires	130
3.3. Performances du système	134

Chapitre IV : Poursuite du travail

4.1. Implémentation du système complet en temps réel	136
4.1.1. Enchaînement avec le niveau acoustique	136
4.1.2. Système temps réel travaillant en mode conversationnel	137
4.2. Généralisation du système	140
4.2.1. Augmentation de la taille du vocabulaire accepté	140
4.2.2. Augmentation du nombre de locuteurs	143
4.3. Phase d'apprentissage	143
4.3.1. au niveau phonémique	143
4.3.2. au niveau morphologique	144
4.3.3. au niveau syntaxique	144

4.4. Autres possibilités	145
4.4.1. Caractérisation sémantique d'un langage	145
4.4.2. Définition de langages oraux de communication Homme-Machine	146

CONCLUSION

BIBLIOGRAPHIE

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

La reconnaissance automatique de la parole s'inscrit dans le cadre général des recherches sur le dialogue entre l'homme et la machine. Née de l'informatique et largement tributaire de ses moyens, cette science n'en possède pas moins ses caractères propres dont le principal est la pluridisciplinarité. Elle amène ainsi le chercheur à aborder des domaines aussi divers que l'acoustique, l'automatique, l'informatique, la linguistique, la phonétique...

Les réalisations actuelles restent encore modestes et les solutions proposées très partielles. Mais il ne faut pas voir à travers cela un constat d'échec : la reconnaissance automatique de la parole est un problème très complexe dû à la particularité du signal vocal. Parmi les réalisations proposées jusqu'alors, on peut noter deux approches différentes : la reconnaissance de mots isolés qui, comme nous le verrons, a été la plus étudiée et dont on commence actuellement à connaître les principes et la reconnaissance de phrases continues dont on ne sait encore que très peu de choses et pour laquelle peu de résultats ont été présentés jusqu'alors. De plus, on distingue deux types de systèmes : les systèmes à reconnaître la parole (Speech recognition systems) et les systèmes à comprendre la parole (Speech understanding systems). Ces derniers utilisent le plus souvent des techniques analogues à celles développées en intelligence artificielle.

Ce mémoire présente un système de reconnaissance et de compréhension du discours continu fondé sur un analyseur syntaxique. Nous avons voulu le faire de manière générale, ce qui nous a conduit à considérer le langage reconnu comme un paramètre du système. Il ne s'agit pas là d'un aboutissement final mais plutôt d'une étape vers la reconnaissance de langages plus vastes tels que les langues naturelles. Les résultats obtenus dans le cadre d'un langage particulier sont déjà intéressants et semblent prometteurs pour la poursuite des recherches.

Cette étude comporte trois parties :

La première partie rassemble quelques notions sur l'étude du signal vocal nécessaires à la compréhension de la suite du travail. On y trouve aussi une présentation générale de la reconnaissance de la parole ainsi qu'un bilan des recherches faites en ce domaine.

La seconde partie comporte la présentation générale du système que nous avons mis au point. Les niveaux syntaxique et morphologique y sont plus spécialement détaillés ainsi qu'une méthode permettant de garder une certaine liberté par rapport à la grammaire du langage traité.

La troisième partie, enfin, présente l'application de ce système à un langage particulier, le langage choisi étant un langage de communication avec un standard téléphonique. L'importance d'une phase de dialogue, propre à chaque application, y est plus spécialement étudiée : c'est elle, en effet qui permet le plus souvent de valider la réponse fournie par le système de reconnaissance. L'ensemble du système a été testé sur le CII 10070 de l'IUCA de NANCY, en temps différé ; les résultats sont présentés à la fin de ce mémoire ainsi que les grandes lignes de la poursuite de ce travail.

P A R T I E A

LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE

INTRODUCTION ET
ENONCES DES PROBLEMES.

INTRODUCTION A LA PARTIE A

Cette première partie est destinée à situer le problème de la reconnaissance automatique de la parole et à introduire un certain nombre de notions nécessaires pour la compréhension du travail effectué.

De par sa nature même la reconnaissance vocale utilise des disciplines aussi variées que l'acoustique, la phonétique, la physiologie, l'informatique.... On trouvera ainsi des notions sommaires sur le signal vocal : ses caractéristiques et les sons élémentaires qui le composent.

Puis, après avoir défini la reconnaissance de la parole, ses buts et ses limitations, une revue des différentes approches utilisées et des travaux antérieurs est proposée. Cette dernière est surtout tournée vers les systèmes utilisant des contraintes de hauts niveaux : syntaxiques, sémantiques, lexicographiques.., qui nous intéressent plus spécialement.

Enfin une présentation de l'utilité des entrées vocales termine cette partie.

C H A P I T R E I

INTRODUCTION A L'ETUDE DE LA PAROLE



1-1- GENERALITES :

La parole est sans aucun doute le moyen le plus utilisé pour faire transiter des informations d'un individu à un autre. Ce n'est certes pas le moyen le plus fiable mais cet inconvénient est largement compensé par la facilité d'utilisation.

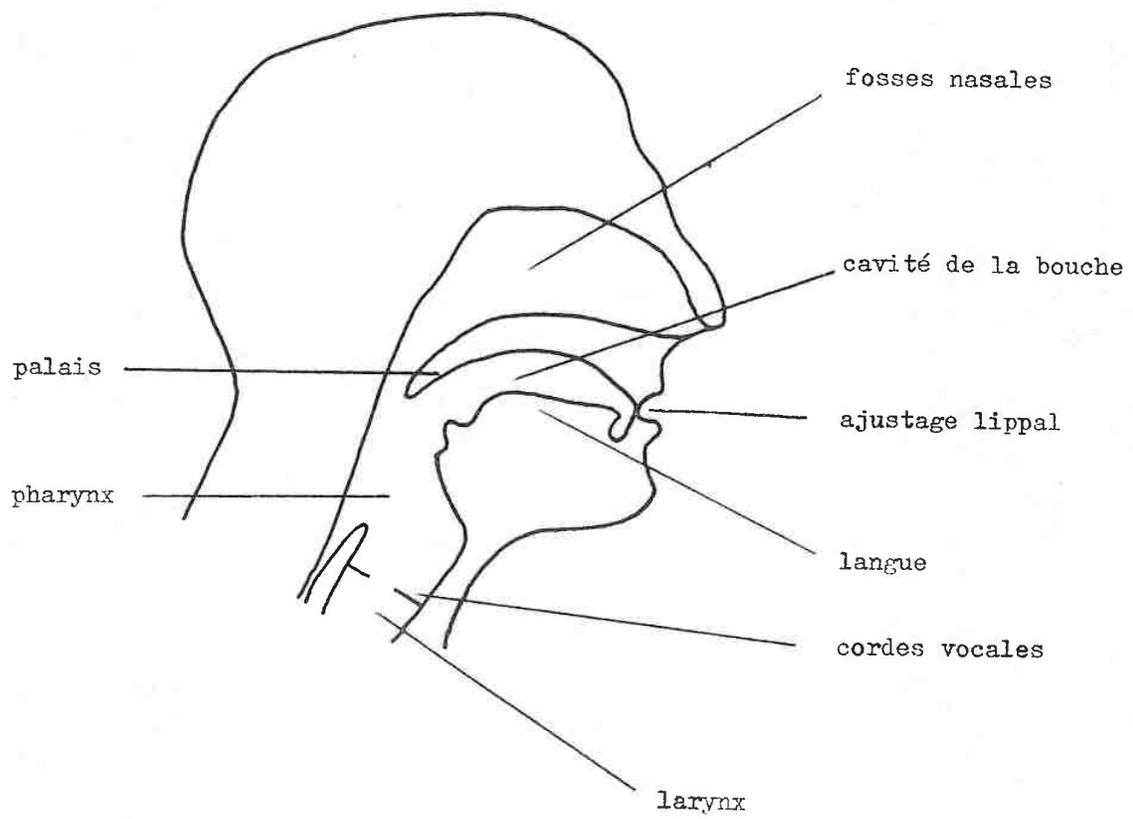
A un moment où s'effectue de plus en plus de recherches sur la communication homme-machine (systèmes conversationnels, aide à la programmation...) il n'est pas étonnant que la parole, sa reconnaissance et sa synthèse, soit un pôle important des travaux de recherches. Avant de voir plus précisément les problèmes liés à la reconnaissance, et bien que l'on tienne en général pour connus les phénomènes d'émission et de perception de la parole, il nous a paru nécessaire de rassembler ici quelques notions fondamentales sans lesquelles on peut difficilement parler de reconnaissance de la parole.

Sans vouloir présenter tous les éléments nécessaires à ce travail, nous étudierons quelques caractéristiques de la parole utiles pour comprendre les mécanismes d'émission et de perception d'un langage(1).

1-2- LE SIGNAL DE LA PAROLE :

1-2-1- Définition :

La parole est un phénomène vibratoire résultant de deux composantes : d'une part une source d'énergie acoustique, les cordes vocales, d'autre part un système résonant, le canal vocal, composé de plusieurs cavités:



Appareil humain d'émission de la parole

Figure 1.1.

le pharynx, la cavité de la bouche, les fosses nasales. (La figure 1.1. donne un schéma général de l'appareil phonatoire humain). Son étude se fait en général à l'aide de sa forme électrique obtenue par un microphone qui transforme les variations de pression en variations de courant.

1-2-2- Caractéristiques du signal vocal :

Par rapport aux autres signaux utilisés pour la communication entre l'homme et la machine, la parole est un signal à part qui a plusieurs caractéristiques particulières cachant chacune une difficulté à surmonter lors de de la reconnaissance.

Ces différents caractères sont :

a) la continuité : à la différence de l'écriture où les mots sont séparés les uns des autres par des blancs, le langage oral est une suite pseudo-continue de sons. Les arrêts correspondent en général à des pauses de respiration ou à des occlusions précédant les consonnes plosives et il n'y a aucune correspondance systématique entre ces arrêts et les blancs du langage écrit. Nous retrouverons cette difficulté de séparer les différents mots de la phrase prononcée lorsque nous étudierons l'analyse lexicographique (Partie B, Chapitre 4).

b) la variabilité : dans le cas d'un locuteur unique, plusieurs différences de prononciation apparaissent suivant que l'on crie, parle ou chuchote, ou suivant que l'on fait des liaisons ou non. Nous verrons (Partie B, Chapitre 2) qu'il nous faudra tenir compte de cette variabilité dans la représentation de nos informations, spécialement dans le lexique.

Le problème se complique lorsqu'il faut traiter des messages émis par plusieurs locuteurs. Il y a de nettes différences entre des voix d'hommes, de femmes ou d'enfants et, à l'intérieur de chacune de ces trois classes, d'autres différences apparaissent suivant l'accent (le signal vocal varie énormément pour une même phrase suivant qu'elle est prononcée par un nancéien, un toulousain ou un parisien et ces différences sont encore plus nettes lorsqu'un étranger s'exprime en français).

Dans la suite de ce travail nous nous limiterons à un seul locuteur (voix parlée), la généralisation à plusieurs locuteurs devant faire l'objet d'une étude ultérieure.

c) la redondance : dans la parole deux types d'informations sont transmises : une information sémantique qui contient le sens du message et qui est seule nécessaire à la compréhension et une information parfois qualifiée d'esthétique qui caractérise le locuteur (accent, ton de la voix...). On se rend particulièrement compte de cette redondance quand on constate que 50 bits par seconde suffisent à transmettre la parole sous forme de suite de phonèmes, alors qu'il faut 30 000 bits par seconde pour la transmission téléphonique et 200 000 bits par seconde pour la "haute fidélité".

Cette énorme redondance accroît l'intelligibilité de la parole. Elle permet ainsi une bonne compréhension d'un message, même émis au milieu de bruits parasites importants, malgré la grande variabilité du signal de parole et elle renforce la résistance de l'information sémantique d'un message parlé aux diverses perturbations dues à l'environnement. Les méthodes d'analyse que nous présentons au chapitre 2 ont pour but de diminuer cette redondance sans perdre, autant que possible, d'informations utiles.

d) le dernier caractère qu'il est bon de noter est la très grande liberté qu'offre le langage parlé. La syntaxe y est souvent plus difficilement respectée que dans le langage écrit. Il faudra donc trouver des outils permettant d'obtenir une certaine liberté par rapport à la syntaxe sans pour autant trop alourdir le système de reconnaissance. Nous proposerons une solution partielle à ce problème dans l'étude du vérificateur d'hypothèses (Partie B, Chapitre 4).

1-3- SONS ELEMENTAIRES DE LA PAROLE :

1-3-1- Critères de classement des sons du langage :

Ce classement des sons du langage se fait principalement à partir de trois critères (2) :

a) le voisement provoqué par l'activité du larynx (cordes vocales),

b) le mode d'articulation ou type du mécanisme,

c) le point d'articulation, c'est-à-dire l'endroit de resserrement maximal du conduit vocal.

Ce classement des phonéticiens apparaît insuffisant en reconnaissance car il ne tient pas toujours compte des phénomènes acoustiques (phénomènes de compensation) (3) .

L'unité minimale d'une langue est en effet le phonème mais la définition linguistique du phonème est inutilisable car les traits utilisés pour les diverses classifications sont difficiles à reconnaître en pratique. On utilisera donc une définition restrictive du phonème par commodité : dans la suite on appellera phonème un son résultat de la simultanéité dans le temps d'un ensemble de traits acoustiques et éventuellement phonétiques.

Nous distinguerons ainsi trois types de sons : les voyelles, les consonnes et l'ensemble formé des liquides et des semi-voyelles. Les premières correspondent à une configuration relativement stable du conduit vocal par rapport aux secondes. Quant aux troisièmes, voisines des voyelles par leur mode d'articulation, elles se caractérisent par un moins bon rendement acoustique.

1-3-2- Les voyelles :

On différencie en français deux classes de voyelles :

a) les voyelles orales, au nombre de douze, émises exclusivement par la bouche :

/a/, /à/, /i/, /e/, /é/, /ə/, /ø/, /œ/, /ɔ/, /o/, /u/, /y/

b) les voyelles nasales au nombre de quatre :

/ô/, /â/, /ê/, /œ/

Lors de la reconnaissance, on se limitera en fait à moins de voyelles que cela : Il ne sera pas fait de distinctions entre o et ə, a et â... car les différences de prononciation de ces voyelles sont de moins en moins pertinentes dans la langue courante.

1-3-3- Les consonnes :

Il existe des consonnes sourdes et des consonnes sonores, selon que le flux d'air est modulé ou non par les cordes vocales.

Les consonnes sont classées en deux grands types :

a) les occlusives : elles correspondent à une fermeture du canal suivie d'une ouverture brusque. Les occlu-

sives françaises sont :

$/p/$, $/t/$, $/k/$ sourdes

$/b/$, $/d/$, $/g/$ sonores

b) les fricatives : elles sont caractérisées par une sorte de sifflement dû au passage de l'air dans une partie rétrécie du conduit vocal. Les fricatives françaises sont :

$/f/$, $/s/$, $/ʃ/$ sourdes

$/v/$, $/z/$, $/ʒ/$ sonores

1-3-4- Les liquides et semi-voyelles :

Leur mode d'articulation est voisin de celui des voyelles mais le conduit vocal étant moins ouvert, elles ont un moins bon rendement acoustique, et l'énergie globale du signal résultant est plus faible.

On distingue en français les liquides $/l/$ et $/r/$
et les semi-voyelles $/w/$ et $/j/$

C H A P I T R E I I

LA RECONNAISSANCE AUTOMATIQUE DE LA PAROLE



2-1- DEFINITION :

Une machine à reconnaître la parole peut être définie comme un système capable de comprendre un énoncé oral : son, mot, message parlé... Il s'agit, grâce à diverses transformations du signal vocal, de faire entrer dans le domaine de connaissance de l'ordinateur le contexte d'un échange oral. Le but de la reconnaissance est de transformer la parole en une action. Nous entendons par action soit l'écriture du message émis, soit l'exécution de l'ordre ou de la demande contenue dans le message.

2-2- METHODES D'ANALYSE DU SIGNAL ET PARAMETRISATION :

Sans faire un exposé complet sur les méthodes d'analyse ou de paramétrisation applicable à la parole, nous nous contenterons de quelques rappels utiles pour la compréhension du système complet de reconnaissance.

Dans ce qui suit, nous désignerons le signal vocal en un point donné par la fonction $f(t)$: elle représente les variations de la pression acoustique en ce point dues à l'onde sonore (on l'étudie le plus souvent sous la forme électrique obtenue à l'aide d'un microphone).

2-2-1- Méthodes spectrales ou fréquentielles :

La représentation de la parole dans le domaine fréquentiel présente un certain nombre d'avantages : il est en effet possible de décrire les sons de la parole en terme de fréquence à partir du mécanisme d'émission vocale.

Mais la parole étant un phénomène non stationnaire, où le temps joue un rôle primordial, il faut donc obtenir un spectre évolutif permettant d'introduire la dimension temporelle.

a) analyse par transformée de Fourier :

Un calcul numérique peut nous fournir la transformée discrète de Fourier (D.F.T.) de $f(t)$. On définit ainsi un spectre de façon discrète : à N échantillons x de $f(t)$ correspondent N valeurs X du spectre calculables par ordinateur. Mais les relations définissant la DFT nécessitent des temps de calcul beaucoup trop longs, on utilise donc en fait un algorithme de transformée rapide de Fourier (FFT). Cet algorithme proposé initialement par Cooley et Tukey (4) permet de calculer la DFT avec une telle réduction du temps de calcul qu'il est possible d'obtenir une transformée de Fourier d'un signal oral en temps réel à l'aide d'un gros ordinateur ou avec un processeur câblé.

b) analyse par bancs de filtre :

Cette méthode permet d'obtenir un spectre à court terme $|F(\omega, t)|$ autour de la fréquence $\omega/2\pi$ en filtrant $f(t)$ par un filtre passe-bande calé sur cette fréquence, puis en effectuant un redressement et un filtrage passe-bas du signal obtenu. Ainsi, à l'instant t pour une fréquence $\omega/2\pi$ on a l'amplitude correspondante du spectre à court terme. En utilisant plusieurs circuits identiques correspondants à des fréquences différentes, on obtient la représentation du spectre à court terme dans un système à trois dimensions (temps, fréquence, amplitude). On peut transformer ce spectre dans un système à deux dimensions en projetant l'amplitude sur le plan temps-fréquence : ceci peut être représenté sur un écran de visualisation par un noircissement plus ou moins grand suivant l'amplitude.

Une variante de cette méthode fournit une analyse spectrale binaire (1) : on ne fait plus intervenir l'amplitude que sous une forme binaire, 0 ou 1. Pour chaque paire de filtres consécutifs F_n et F_{n-1} on note 0 si la tension de sortie du filtre F_{n-1} est supérieure à celle du filtre F_n et 1 dans le cas contraire.

c) autres méthodes spectrales :

Il existe bien d'autres méthodes d'analyse spectrale, on peut citer parmi les plus importants l'analyse par corrélation, l'analyse par transformée de Walsh facilement implémentée sur petit ordinateur mais donnant moins d'informations que la transformée de Fourier. Cette méthode n'apporte donc rien de plus dans le cas du traitement de la parole.

2-2-2- Méthodes temporelles :

D'autres méthodes d'analyse de la parole se basent sur le domaine temporel plutôt que sur le domaine fréquentiel. Un type très simple d'analyse temporelle consiste à compter le nombre de passages par zéro ("zéro-crossings") du signal de parole. L'un des premiers à appliquer cette méthode à la parole fut Peterson en 1951 (5). L'intérêt réside dans le fait que le taux de zéro-crossings du signal de la parole et sa dérivée correspondent approximativement au premier et second formants, qui sont des paramètres intéressants en reconnaissance vocale et c'est d'autant plus séduisant que l'on peut facilement les obtenir par un dispositif électronique (6). Mais ces paramètres, présentant une trop grande variabilité, ne peuvent en fait être utilisés que comme paramètres secondaires.

Une autre méthode intéressante est la méthode d'analyse par prédiction linéaire du signal vocal (7). Ce "codage prédictif" est la seule méthode utilisant le caractère anticipatif de la parole; il part du principe que l'on peut prédire approximativement la valeur d'un échantillon de parole à partir d'un développement limité des échantillons précédents.

2-2-3- Détermination de paramètres :

Déterminer des paramètres de la parole c'est extraire un nombre limité de caractères pertinents à partir du signal vocal ou de données fournies par une analyse préliminaire. Il s'agit d'effectuer une compression d'information, soit par un dispositif câblé, soit par un algorithme. On peut dire qu'analyse et paramétrisation ne font qu'un, on distingue pourtant la seconde opération qui fournit des paramètres en nombre plus restreint mais plus élaborés.

Comme dans toute étude de reconnaissance des formes, le problème de paramétrisation est fondamental en reconnaissance de la parole, car le nombre de paramètres et leur qualité déterminent la rapidité et la performance du processus de reconnaissance.

On distingue deux types de paramètres : les paramètres physiques et les paramètres mathématiques.

a) les paramètres physiques :

Ces paramètres ont toujours un caractère plus ou moins subjectif; les deux plus importants et plus utilisés en reconnaissance de la parole sont la fréquence fondamentale et les fréquences des formants. La recherche du fondamental de la voix a fait l'objet d'un certain nombre d'études, bien qu'on ne possède pas encore de solution entièrement satisfaisante. L'extraction des formants a été beaucoup plus étudiée car ces paramètres sont essentiels pour la caractérisation des voyelles ; le codage prédictif, sous diverses formes fournissant actuellement les méthodes les plus sûres mais encore imparfaites.

b) les paramètres mathématiques :

La paramétrisation peut aussi se faire par une opération mathématique :

- de compression d'information fournissant une représentation de la parole dans une certaine base. Diverses méthodes de compression d'information ont été utilisées, elles donnent toutes des résultats du même ordre (8) et sont plus ou moins bien adaptées à chaque problème.
- de traitement du signal : obtention des coefficients de prédiction par une méthode de codage prédictif.

2-3- DIFFERENTES APPROCHES DE LA RECONNAISSANCE :

Il existe en reconnaissance de la parole deux approches bien différentes : la reconnaissance globale
la reconnaissance analytique.

Elles se différencient l'une de l'autre tant par leur principe que par le domaine d'application qu'elles ouvrent.

2-3-1- Reconnaissance globale :

Cette approche, appelée aussi reconnaissance par "mots" ("mots" signifiant ici tant des mots au sens courant que des expressions ou même des phrases), calque à peu près le fonctionnement du cerveau lorsqu'il comprend un énoncé oral dans son ensemble, sans s'arrêter aux différents composants du message. Il s'agit de comparer globalement le "mot" aux différents mots de référence stockés dans un dictionnaire. Le traitement acoustique préliminaire est

assez simplifié; il n'y a pas de problème de segmentation, par contre, il faut conserver en mémoire les différentes représentations possibles de tous les mots.

Son principal inconvénient lors du traitement de phrases réside dans le fait que les mots doivent être séparés les uns des autres par un silence (de l'ordre d'une demi-seconde). Cela nécessite un énoncé mot par mot des phrases et cette contrainte, difficilement mise en place dans le cas de langages se rapprochant de langues naturelles (français par exemple), donne une impression très artificielle au dialogue ainsi obtenu.

Cette approche a été jusqu'ici la plus utilisée ; elle nécessite une normalisation de la parole en fréquence et en temps et se satisfait d'une analyse assez rudimentaire du signal vocal, car elle se fonde plus sur l'aspect global d'un mot que sur sa structure fine. Les études faites sur cette approche ont montré sa validité pour des vocabulaires de taille moyenne, dont la limite doit se situer aux environs de quelques centaines de mots.

2-3-2- Reconnaissance analytique :

Cette approche, plus générale, permet de traiter des vocabulaires pratiquement illimités ; de plus c'est la seule utilisable en reconnaissance du discours continu. Les résultats obtenus sont moins spectaculaires que dans le cas de la reconnaissance globale, car d'importants problèmes se posent tels que la segmentation du signal vocal. En effet, elle travaille sur des sons élémentaires (phonèmes, diphonèmes, syllabes, etc...) et la principale difficulté est d'obtenir la suite d'éléments phonétiques correspondant à l'énoncé oral de départ. Le processus nécessite donc dans ce cas différentes étapes: la segmentation, l'identification des segments et la reconnaissance de la phrase.

Comme nous l'avons déjà noté cette approche est la seule qui permette d'aborder le problème du discours continu en association avec des analyses au niveau lexical, linguistique, syntaxique et sémantique. C'est donc dans le cadre d'une reconnaissance analytique que nous avons travaillé à la mise au point de notre système, en nous limitant aux niveaux lexical, syntaxique et sémantique.

2-4- TRAVAUX ANTERIEURS SUR LA RECONNAISSANCE DE LA PAROLE :

2-4-1- Systèmes électroniques :

Les premiers systèmes de reconnaissance de la parole ont été proposés vers 1950 sous forme d'appareillages électroniques. Bien que très limités, ces appareils ont cependant contribué à faire avancer nos connaissances dans ce domaine.

La première réalisation fut, sans doute, celle de J. DREYFUS-CRAF en 1950 (9) . Cet appareil, nommé phonéto-
graphe fournissait une représentation graphique des sons étudiés. DREYFUS-CRAF a produit différentes versions améliorées de son appareil, dont la plus récente permet la reconnaissance de 18 mots courts prononcés par plusieurs locuteurs.

Le premier appareil capable de reconnaître les dix chiffres fut présenté en 1952 par les laboratoires BELL (10) . Le taux de reconnaissance atteignait 99 % pour un seul locuteur. En 1958, une version plus évoluée proposée par DUDLEY et BOLOSHEK (11) donnait un taux de reconnaissance de 90 % pour divers locuteurs à condition qu'ils modifient leur voix.

Une autre réalisation intéressante fut présentée en 1956 : la "machine à écrire phonétique" de OLSON et BELON (laboratoires RC.A.) (12) capable de reconnaître dix mots très courts prononcés par un seul locuteur en utilisant la syllabe comme unité de reconnaissance.

En 1958, une nouvelle étape est franchie par FRY et DENES (13) . Ils construisirent un système à deux "étapes" dans lequel une reconnaissance purement acoustique était affinée par l'utilisation de données linguistiques. L'amélioration obtenue étant de 20 %, ils prouvèrent ainsi la validité de la méthode.

2-4-2- Systèmes informatiques :

Les ordinateurs, apparus en ce domaine vers 1958, ont donné une nouvelle dimension aux recherches et ont permis de dépasser les limites de faible capacité de mémoire, caractéristiques des systèmes entièrement électroniques. L'utilisation de l'ordinateur en reconnaissance de la parole devient quasi-générale à partir de cette date.

Les premières expériences d'utilisation de l'ordinateur furent réalisées aux Etats-Unis (1959-1961) dans les laboratoires du M.I.T. et d'I.B.M. Tous ces chercheurs effectuaient une analyse spectrale de la parole à l'aide d'un système Hardware sauf GCHULTZ (I.B.M.) (14) dont le programme, comportant une analyse simulée du signal vocal, permettait la reconnaissance des 10 chiffres à 97 % pour 50 locuteurs, hommes et femmes. Tous concluaient à l'utilité indiscutable de l'ordinateur en reconnaissance de la parole.

Plusieurs types d'analyse ont été utilisées à cette époque : analyse fréquentielle, analyse par "zero-crossing"... Et, tout en restant encore très limités, ces programmes de reconnaissance représentaient déjà un gros progrès puisqu'ils admettaient un nombre plus important de locuteurs.

Au fur et à mesure du développement des ordinateurs des expériences plus importantes furent tentées. Ainsi, vers 1966, on pouvait reconnaître correctement 30 à 50 mots pour plusieurs locuteurs (15) , (16) .

2-5- ETAT ACTUEL DES RECHERCHES :

Actuellement, il reste encore beaucoup à faire en reconnaissance vocale. La reconnaissance globale, en temps réel, d'un vocabulaire restreint est chose à peu près acquise, pour un nombre limité de locuteurs du même sexe. Mais tout le domaine de la reconnaissance analytique reste à étudier spécialement en ce qui concerne la reconnaissance de phrases ou de vocabulaires importants.

2-5-1- Etude sur l'analyse et la paramétrisation de la parole:

En vue d'une reconnaissance analytique du discours continu, plusieurs équipes de chercheurs travaillent sur l'analyse et la paramétrisation de la parole.

A Stanford (USA) REDDY fut l'un des premiers à aborder les problèmes de segmentation et de reconnaissance de la parole continue (17) . Les recherches menées à Cambridge concernent à la fois l'analyse et la reconnaissance. BOBROW et KLATT se sont attachés à trouver des paramètres intéressants pour la reconnaissance dans le cadre d'un système capable de reconnaître 100 mots ou courtes phrases (18).

En URSS, plusieurs équipes ont fait des études sur la recherche de paramètres dans le cadre de la reconnaissance de la parole (18) , (19) .

En France aussi plusieurs équipes s'intéressent à ces problèmes : (20) , (21) , (22) , (23) .

2-5-2- Prise en compte des niveaux morphologique, syntaxique et sémantique :

L'étude de la reconnaissance du discours continu a mis en évidence la nécessité d'utiliser des contraintes de niveaux hiérarchiques supérieurs : morphologique, syntaxique, sémantique, linguistique. Depuis 1969, l'étude de ces divers niveaux dans le cadre de la reconnaissance de la parole a pris une importance croissante. Les Etats-Unis restent en tête dans ce domaine. Mais il ne faut pas négliger les résultats des chercheurs soviétiques. En Europe enfin plusieurs équipes commencent à aborder ce problème et il y a des résultats notables, spécialement en France. Le travail présenté dans ce domaine s'inscrit dans ce cadre.

a) Aux Etats-Unis :

Dès 1968, ALTER et REDDY notaient l'utilité des contraintes contextuelles dans la reconnaissance automatique de la parole (24) , (25) . Ce fut VICENS (6) qui le premier présenta un système utilisant les données fournies par la syntaxe du langage. Il est intéressant de citer quelques résultats : par exemple "92 % de reconnaissance après 16 secondes environ pour une liste de 561 mots ou courtes phrases prononcées par un locuteur". Deux applications sont proposées par VICENS : une machine à calculer de Bureau, DESCAL, et un robot manipulateur, HAND-EYE-EAR, tous deux commandés par un langage à syntaxe très simple.

Mais la plupart des systèmes présentés jusqu'alors et utilisant des contraintes syntaxiques ont surtout travaillé sur des phrases à mots isolés (6) , (26) , (27) . Les travaux sur la reconnaissance du discours continu ont débuté

sérieusement aux USA dans le cadre d'un projet de l'ARPA (rapport Newell).

Actuellement plusieurs systèmes sont en cours d'étude.

Ainsi REDDY travaille à un projet baptisé HEARSAY II qui doit permettre de reconnaître des phrases d'un anglais simplifié à propos d'un sujet particulier (28) , (29) , (30) ; le vocabulaire est d'environ 1200 mots. Ce système, basé sur un principe d' "Hypothèse et Test", utilise une analyse descendante et essaie de reconnaître d'abord les phrases ayant la plus forte probabilité d'apparaître. Ce système est actuellement en cours d'essais.

KLOMSTAT et MONSHEIM élaborent un système (CASPER) (31) où l'utilisateur a la possibilité de définir la grammaire du langage qu'il veut reconnaître. Ils utilisent un analyseur syntaxique qui indique les classes syntaxiques des mots pouvant apparaître dans la phrase et un comparateur de mots qui détermine les mots en accord avec les prévisions syntaxiques et la chaîne d'entrée.

MILLER dans le système LPARS (32) essaie dans un premier temps de reconnaître des mots qu'il considère comme des îlots de confiance. De là, il tente des analyses partielles sur leur droite et sur leur gauche, produisant ainsi des arbres d'analyse qu'un algorithme essaie de rassembler pour faire une analyse syntaxique complète et cohérente. Le vocabulaire utilisé est de 70 mots.

BAKER (33) s'appuie sur un modèle probabiliste, en l'occurrence un processus MARKOV, pour reconnaître la phrase à partir des données acoustiques. Les différents niveaux acoustique, phonologique, syntaxique et sémantique sont, pour lui autant de processus de MARKOV. Cette approche probabiliste, quoique assez lourde (le nombre de chemins possibles est très grand), est très intéressante mais

ce système est en cours de développement et peu de résultats sont disponibles.

Dans le cadre du projet SDC, BARNETT et RITEA ont présenté un système de gestion de donnée utilisant des entrées-sorties orales (V.D.M.S.) (34) , (35) . Ce système est basé sur ce qu'ils appellent les "contraintes linguistiques prédicatives" (PLC). Ces contraintes, de type syntaxique, permettent au système, à l'aide de la grammaire du langage, d'émettre des prédictions sur les mots à reconnaître. Ces prédictions sont alors reprises par une procédure de recherche lexicale dont la tâche se limite simplement à vérifier que le mot prévu est bien dans la chaîne d'entrée. La grammaire testée comporte 35 règles et 150 mots et, lors des tests faits pour deux locuteurs hommes, les résultats obtenus donnent une reconnaissance correcte pour 52 % des phrases prononcées.

Enfin le système mis au point dans le cadre du projet IBM de reconnaissance du discours continu (36) utilise lui aussi des informations syntaxiques en lien avec des informations lexicales et statistiques.

b) en URSS :

Il existe peu de communications soviétiques sur le sujet. Néanmoins les Russes s'intéressent de près à la reconnaissance de la parole comme l'ont montré les derniers congrès spécialisés (ARSO VI et VII) et leur participation au 6e J.E.P. à Toulouse en Mai 1975. L'équipe de VYSOTKIY a noté, lors d'expériences sur la reconnaissance de 50 instructions (37) , l'intérêt des informations linguistiques en reconnaissance de la parole. Il semble que leur recherche soit surtout tournée vers la commande orale de machine (38) .

c) En Europe :

Très peu d'équipes européennes travaillent à ces problèmes et la reconnaissance de la parole continue n'est abordée que depuis peu de temps. En France, une dizaine d'équipes travaillent actuellement sur la reconnaissance de la parole ; parmi elles quelques unes s'intéressent au rôle de la syntaxe ; G. MERCIER et son équipe au CNET de Lannion, BELLISSANT à l'Université de Grenoble ; le groupe d'intelligence artificielle à Aix-Marseille. Des systèmes de reconnaissance par mots isolés ont déjà été présentés. Ce fut le cas de TUBACH (39) qui reconnaît un programme Algol, prononcé par un seul locuteur, quasiment à 100 % , à l'aide de contraintes syntaxico-sémantiques. A Nancy un système de commande machines outils a été mis en place à l'aide d'une analyse syntaxique par J.P. HATON (40) .

Le travail présenté dans ce mémoire est une contribution à l'étude de systèmes capables de comprendre une conversation continue, menée actuellement par notre équipe à Nancy. Nous obtenons déjà des premiers résultats mais, comme nous le verrons en partie C, il reste encore beaucoup de points à travailler avant de pouvoir dominer ces problèmes.

C H A P I T R E I I I



UTILITE DES ENTREES VOCALES EN MACHINE

3-1- INTERET DE LA PAROLE PAR RAPPORT AUX AUTRES MODES DE COMMUNICATION .

La communication homme-machine utilise actuellement des moyens de deux types : manuel (écriture, perforation, interrupteurs etc...) et visuel (console de visualisation, lecture optique...). Sans vouloir en aucun cas faire disparaître les modes de communication déjà utilisés, il semble que la parole doit pouvoir prendre sa place comme moyen privilégié de dialogue.

La parole est en effet le moyen de communication le plus naturel pour l'homme, son usage nous est familier et facile, en ce qui concerne du moins notre langue maternelle. Elle convient bien à l'émission de messages et seule elle permet de le faire lorsqu'on est occupé par ailleurs. Cette caractéristique est importante spécialement dans des situations d'alarme ou de danger où la seule transcription manuelle du message d'alerte peut faire perdre un temps précieux. Elle peut aussi être utilisée simultanément avec les autres moyens de communication.

Sa propagation est omni-dimensionnelle, ce qui permet de ne pas toujours être en face du poste de saisie pour transmettre des informations, et de plus, elle est facilement transformable sous sa forme électrique, avantage certain dans le cas de transmission à longue distance.

Le poste de saisie nécessaire pour la capter est simple et peu encombrant : il suffit d'un micro qui, comme nous l'avons déjà noté, n'est pas nécessairement devant la personne qui transmet le message. Tout le traitement se fait au niveau de l'ordinateur qui peut être loin de ce poste de saisie. Ceci est certainement un avantage dans les milieux industriels, même s'il y a un risque d'altération du signal à cause du niveau sonore du bruit ambiant.

Enfin, la parole permet de véhiculer des informations sur le locuteur tant en ce qui concerne sa condition physique, son état émotionnel que son environnement géographique national ou culturel.

3-2- DIFFERENTS TYPES D'APPLICATIONS DE LA RECONNAISSANCE DE LA PAROLE.

Historiquement, le premier but de la reconnaissance de la parole a été la recherche d'un mode de communication facile et efficace. Comme nous l'avons déjà vu, la reconnaissance automatique d'un message parlé permettrait de réduire d'au moins 1000 à 1 la quantité d'informations à transmettre. Vues les techniques actuelles de télécommunication, ce gain ne présente plus guère d'intérêt, si ce n'est dans le cas de communications à très grande distance (extra-terrestres).

Actuellement l'une des principales applications d'un système de reconnaissance automatique de la parole concerne la communication homme-machine. Elle correspond à l'intérêt de la parole comme moyen de communication que nous avons noté au paragraphe précédent.

Plus précisément, nous pouvons noter des domaines d'application aussi variés que la saisie de données. Dans bien des cas les encodeurs magnétiques pourraient être avantageusement remplacés par une saisie orale. On peut ainsi se demander si une gestion complète d'un magasin ne pourrait pas se faire par voie orale à partir de caissières énonçant simplement le nom des articles achetés par un client (l'expérience a été faite aux USA) et l'ordinateur fournissant la facture et parallèlement gérant les stocks et la comptabilité. Et pourquoi ne pas envisager un système de réservation de places uniquement par téléphone. Il n'est pas sûr que ce genre d'applications soit du domaine du rêve. Peut-être apparaîtront-elles d'ici quelques décennies.

- la gestion et la consultation de base de données. Là aussi, il semble que la parole puisse être un moyen privilégié dans des cas de demande de renseignements.

- la rééducation des mal-entendants.

Cette application fait partie de recherches en cours, spécialement à Nancy (41) : grâce à un moyen très simple de communication avec un ordinateur tel qu'une console de visualisation, on peut en effet utiliser un système conversationnel pour la rééducation des mal-entendants.

- l'enseignement assisté.

- l'apprentissage de langues étrangères.

- enfin, tout le vaste domaine de la commande de processus. Sous cette rubrique, on peut regrouper des applications aussi variées que la transmission d'ordres simples à des machines dont le robot "Hand-Eye-Ear" de VICENS (6) est un des premiers exemples, l'appareil à composer un numéro de téléphone des laboratoires BELL (l'utilisation du téléphone par des handicapés-moteur est donnée comme application possible de ce système) (42), la commande plus élaborée de machines. (1) .

L'application abordée en C (automatisation d'un standard téléphonique) relève à la fois de la commande de processus et de la consultation de fichiers.

En fait toutes ces applications possibles de la reconnaissance automatique de la parole dépendent étroitement de la complexité et de la généralité de ce système (taille du vocabulaire, nombre de locuteurs). La plupart demande en effet un système très évolué, acceptant sous forme de parole continue des messages émanant d'un grand nombre de locuteurs et il est peu probable que de tels systèmes voient le jour avant 1980. C'est pourquoi on ne peut considérer un travail tel que celui

présenté dans ce mémoire comme un aboutissement, mais simplement comme un pas de plus vers l'élaboration de systèmes plus complets permettant de donner à la parole sa place spécifique comme moyen de communication avec une machine.

P A R T I E B

UN SYSTEME DE RECONNAISSANCE DU DISCOURS CONTINU

BASE SUR UN ANALYSEUR SYNTAXIQUE.

INTRODUCTION A LA PARTIE B

Cette partie présente le système paramétrable de reconnaissance de la parole que nous avons mis au point. Après des rappels sur l'étage acoustique, la définition des données et les résultats souhaités, les niveaux syntaxique et lexicographique sont plus spécialement étudiés.

On y distingue :

- Un analyseur syntaxique dont la technique assez différente de celle utilisée en compilation est présentée. Il utilise en paramètre la définition du langage donnée sous forme de grammaire et son but est de guider la reconnaissance au moyen d'hypothèses qu'il émet,

- Un analyseur lexicographique qui vérifie les hypothèses émises par le niveau syntaxique. Il utilise de plus, une méthode permettant d'obtenir une assez grande souplesse dans l'énoncé des phrases en acceptant une certaine liberté par rapport à la grammaire,

- Un module de reconnaissance de mots également présenté après une étude des différentes difficultés rencontrées en recherche lexicale et des solutions proposées dans d'autres travaux.

On étudie enfin les résultats fournis par un tel système et l'apport que peut apporter une phase de dialogue pour une meilleure reconnaissance.

C H A P I T R E I

PRESENTATION GENERALE DU SYSTEME

1-1- PLACE DE CETTE REALISATION DANS UN SYSTEME COMPLET DE RE-
CONNAISSANCE.

Dans un système général de reconnaissance de la parole utilisant des contraintes syntaxiques, on distingue deux grandes parties.

a) Un système acoustico-phonétique correspondant à un premier traitement du signal vocal et dont les résultats peuvent être essentiellement de deux types :

- soit une représentation paramétrée de la phrase prononcée (sous forme de spectres ou de coefficients de prédiction),
- soit une représentation phonétique (sous forme de chaînes de syllabes ou de phonèmes).

b) Un système syntaxico-sémantique qui, à partir de données fournies par le système acoustique et celles provenant de la définition syntaxique et sémantique du langage, détermine la signification du message oral.

La figure B-1-1- donne un schéma très général d'un tel système de reconnaissance.

La présente étude s'attache principalement à définir le niveau syntaxico-sémantique. Seuls quelques rappels sur le système acoustique seront donnés au paragraphe B-1-3. En fait, l'important est de pouvoir caractériser les résultats fournis par le traitement acoustique. On se référera, pour cela aux résultats obtenus par J.P. HATON (1) .

Il serait sans doute intéressant de prévoir une interaction entre ces deux étages. Mais ce problème, encore peu étudié, demeure l'un des plus difficiles et ne sera qu'effleuré dans ce travail.

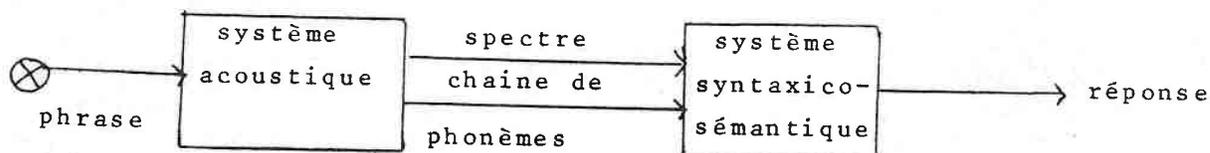


Schéma général d'un système de reconnaissance.

Figure B-1-1-

1-2- POSITION DU PROBLEME :

Pour mieux comprendre l'importance du traitement syntaxico-sémantique, il suffit de réfléchir au mécanisme humain de compréhension.

L'influence de la syntaxe, de la morphologie et de la sémantique existe même si on ne s'en rend pas compte dans le cas de sa langue maternelle, car c'est un fait acquis, passé peu à peu au niveau de l'inconscient. Pour s'en rendre compte, il suffit de remarquer qu'une phrase syntaxiquement incorrecte nous choque, ou même parfois rend impossible la compréhension. C'est encore plus évident dans le cas d'une langue étrangère. Ainsi, si nous prenons le cas d'une personne qui doit dialoguer avec un étranger dont elle ne connaît qu'assez peu la langue, son aptitude à comprendre son interlocuteur sera fortement conditionnée par les connaissances qu'elle a de cette langue au niveau syntaxique, morphologique et sémantique. De plus, lors d'une conversation sur un sujet précis, l'action de la sémantique permet souvent de comprendre une phrase avant sa fin.

Il en est un peu de même au niveau de la reconnaissance de la parole. En fait, comme nous l'avons vu partie A chapitre 2, peu de réalisations ont, pour l'instant été faites dans ce sens, les principaux pôles de recherche se trouvant aux Etats-Unis, spécialement dans le cadre du projet ARPA.

C'est aussi dans ce cadre que se place notre travail. Le système mis au point utilise les informations syntaxiques, morphologiques et sémantiques et se trouve facilement adaptable à divers langages, la définition du langage étant considérée comme un de ses paramètres.

Avant de définir les différents composants de ce système nous donnons quelques rappels sur le système acoustique et les résultats qu'il fournit.

1-3- LE SYSTEME ACOUSTIQUE (43) .

1-3-1- Principe général :

Le système acoustique dont nous utiliserons les résultats peut se diviser en trois parties :

a) La segmentation : elle a pour but de découper le signal vocal en segments minimaux, et consiste à étudier la stabilité du spectre de la parole et à en déduire la position des zones stables et des zones transitoires ; les premières correspondent en général aux corps de voyelles, de fricatives et aux occlusions, alors que les secondes marquent plutôt l'explosion d'une consonne plosive, la fin ou le début d'une voyelle... Les segments ainsi obtenus s'apparentent aux phonèmes, mais il n'y a pas de relations strictes entre segments et phonèmes : une consonne plosive sera en général découpée en deux segments (une occlusion et une explosion) et une voyelle tenue pourra donner lieu à plusieurs segmentations. Ces différents segments sont de plus classés dans quatre grands groupes suivant leur type : fricatif, occlusion de plosive, voyelle, nasal.

b) La reconnaissance : elle a lieu comme la segmentation au cours de la prononciation de la phrase et permet ainsi d'avoir la chaîne phonétique représentant la phrase dès la fin

phoneme
 ↘
 reconnu
 prononcé

	/s/	/ʃ/	/v/	/ʒ/	/z/	/p/	/t/	/k/	/b/	/a/	/o/	/e/	/ə/	/u/	/i/	/â/	/ê/	/r/
/s/	90	6		2	2													
/ʃ/	3	90		6	1													
/v/	10		32			15	19	18	6									
/ʒ/	5	40		55														
/z/	29		10	10	51													
/p/						40	19	18	23									
/t/						9	41	43	7									
/k/						5	42	45	8									
/b/						36	17	15	32									
/a/										85		5	10					
/o/											92		8					
/e/												90	2	8				
/ə/										8		1	89			2		
/u/										1	11			81		7		
/i/												7			88			
/â/											2		4			83	11	
/ê/												6				4	90	
/r/							2	3	2	4						6	8	40
/ʁ/										3		16	7					32

Tableau B1-2

Matrice de confusion de phonèmes

de la prononciation de celle-ci. Dès la détection d'un segment, une classification de ce segment est immédiatement opérée par comparaisons avec des segments types avant l'acquisition du segment suivant. Pour ce faire, le principal problème est celui de la normalisation temporelle : les variations de longueur constatées au niveau des segments peuvent être, en effet, très importantes. Ce problème est résolu par une méthode de comparaison dynamique semblable à celle utilisée en reconnaissance globale de mots (44) , (45) .

c) Lissage de la chaîne phonémique : Sans parler de l'optimisation au niveau linguistique (46), il s'agit ici d'un simple lissage permettant de supprimer ou de diminuer la trop grande redondance de la chaîne phonémique obtenue. Les critères utilisés sont soit purement physiques : suppression des segments inférieurs à 0.1 s, soit morphologiques : concaténation de segments fricatifs successifs s'ils sont de même type, concaténation de segments-voyelles successifs identiques, suppression de queues de phonèmes, non significatives.

1-3-2- Résultats fournis par le système acoustique :

Le résultat fourni par le système acoustique correspond à une transcription phonétique erronée de la phrase prononcée. Les erreurs les plus fréquentes sont d'une part des confusions entre plosives : /p/, /t/, /k/ et /b/, /d/, /g/ , d'autre part des altérations introduites par un effet de seuil de tension dans l'analyse du signal vocal : ainsi /ç/ pourra être certaines fois assimilé à la fricative /s/ ou à un /p/ trop peu intense considéré comme une occlusion de plosive.

Afin de mieux voir les diverses confusions possibles, la figure B-1-2 reprend la matrice de confusion de phonèmes obtenue et publiée par J.P. HATON (1).

Outre ces erreurs de substitution, d'autres erreurs de type insertion ou élision s'ajoutent en fonction de la prononciation de la phrase. Il est à noter que, pour la suite de cette étude, on prendra comme référence les résultats obtenus à NANCY en 1973. L'évolution qui s'est produite au niveau reconnaissance phonémique peut sembler nous mettre en position défavorable, mais un système capable de reconnaître des chaînes phonémiques d'une certaine qualité est assuré de pouvoir traiter, avec encore plus de chances de réussite, des chaînes de qualité supérieure dont on peut espérer avoir bientôt les premiers résultats. De plus, pour augmenter les chances, on autorise les réponses multiples du système acoustique limitées soit à un nombre fixe soit à un nombre variable déterminé en fonction de leur seuil de reconnaissance.

1-4- STRUCTURE GENERALE DU SYSTEME :

1-4-1- Idées directrices :

La réalisation de ce système de reconnaissance a été guidée par divers objectifs dont voici les principaux :

a) le point de départ de cette étude a été diverses considérations sur l'importance du niveau syntaxique en compréhension (cf paragraphe B-1-2). Notre premier objectif a donc été de créer un système où l'analyse syntaxique est considérée comme le guide du processus de reconnaissance. Son principal rôle est d'émettre des hypothèses sur les mots à reconnaître afin d'en vérifier la validité.

a) Le second objectif, conséquence directe du premier est de ne considérer la reconnaissance proprement dite que comme un moyen de vérification des hypothèses émises par le niveau syntaxique. De plus, nous avons considéré que ce vérificateur devait permettre d'obtenir une certaine liberté par rapport à la syntaxe du langage, spécialement en cas d'adjonction ou de suppression d'un mot, si du moins ce dernier n'est pas sémantiquement caractéristique.

c) Comme nous l'avons déjà vu, nous n'avons pas voulu créer un système lié à un langage ou à une application spécifique : cela nous a amené à considérer la définition du langage (syntaxe, morphologie et sémantique) comme une donnée ou un paramètre du système réalisé. L'exemple de langage qui se trouve en partie C n'a servi dans ce sens qu'à valider notre approche, et d'autres langages devront être testés dans un proche avenir.

d) Un autre souci qui a présidé à cette réalisation fut celui de la facilité d'évolution et de maintenance du produit obtenu. A cette fin, le système a été conçu de manière assez structurée et modulaire. Par exemple, les entrées-sorties sont nettement séparées, ce qui devrait faciliter une adaptation à une présentation différente des données et des résultats. Les différentes parties du système sont bien délimitées. Cela permet de modifier un point précis sans pour autant rendre inutilisable le reste.

e) Quant à la réalisation, nous avons voulu obtenir un système facilement adaptable à différents types ou marques de calculateur. Ceci nous a conduit à utiliser un des langages les plus répandus : FORTRAN, assez bien adapté, d'ailleurs, à ce type de problème.

f) Un autre objectif a été la possibilité d'accès à différents résultats intermédiaires afin de rendre possible des études ultérieures sur ces derniers. C'est principalement le cas pour étudier le cheminement suivi au cours de la reconnaissance de mot ou de l'analyse syntaxique.

g) Nous avons enfin considéré qu'il était nécessaire d'adjoindre à l'ensemble une procédure de dialogue qui a la charge de valider en dernier recours les résultats fournis par la reconnaissance, de lever les ambiguïtés ou de corriger certaines erreurs. Ce dialogue est, en fait, un début de traitement sémantique, et est rendu nécessaire, ne serait ce que par sécurité, spécialement dans le cas de commande de machine.

1-4-2- Schéma général du système :

La figure B-1-3 donne un schéma général du fonctionnement du système complet.

On peut y noter 6 grandes parties :

- I- Définition syntaxique et sémantique du langage
- II- Système acoustique et saisie de données
- III- Dictionnaire des mots du langage
- IV- Vérificateur acoustique des Hypothèses
- V- Analyse syntaxique
- VI- Procédure de dialogue.

Nous allons dans la suite étudier ces différentes parties à l'exclusion du système acoustique et de la saisie des données dont les résultats seront considérés comme acquis.

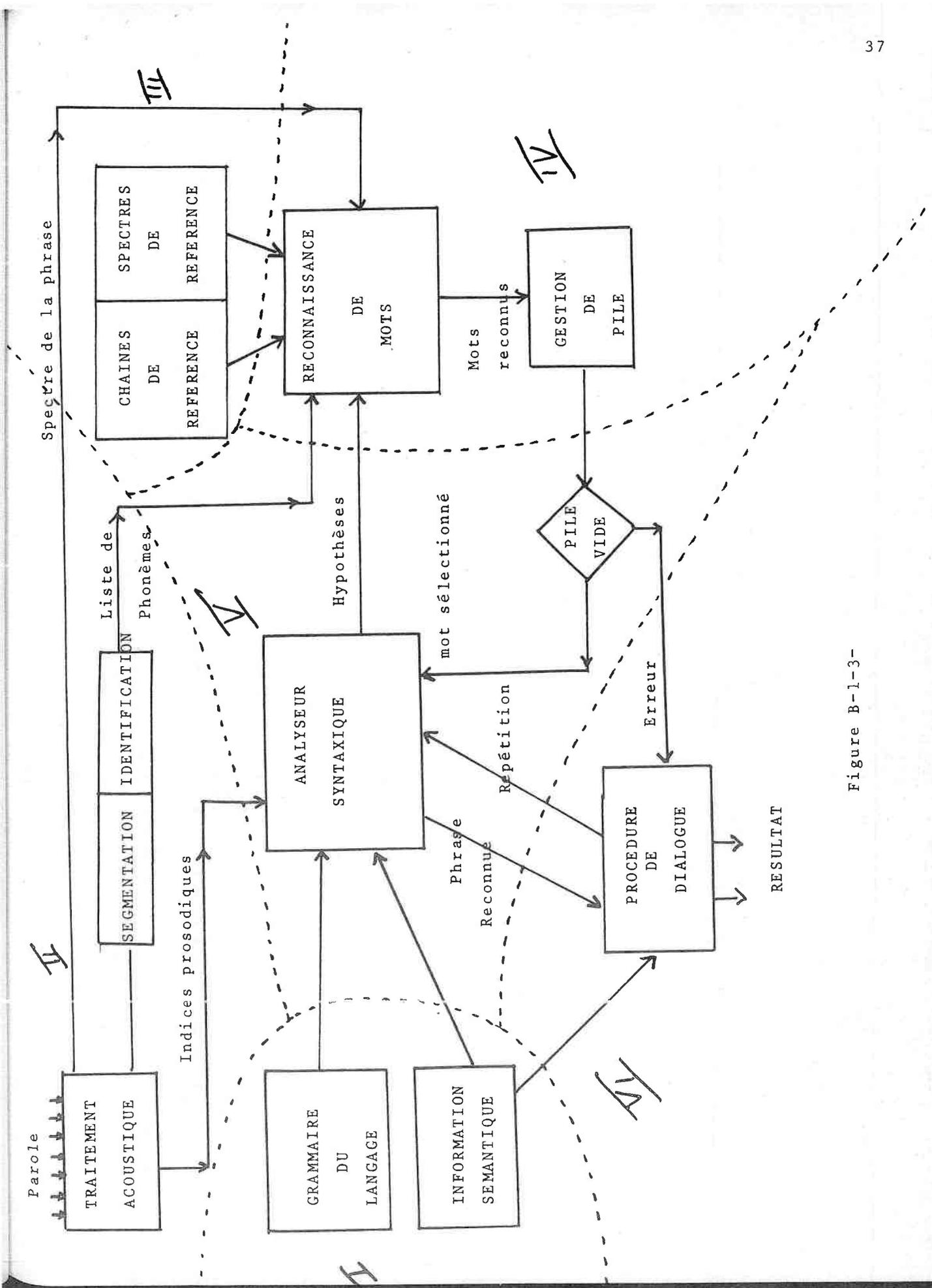
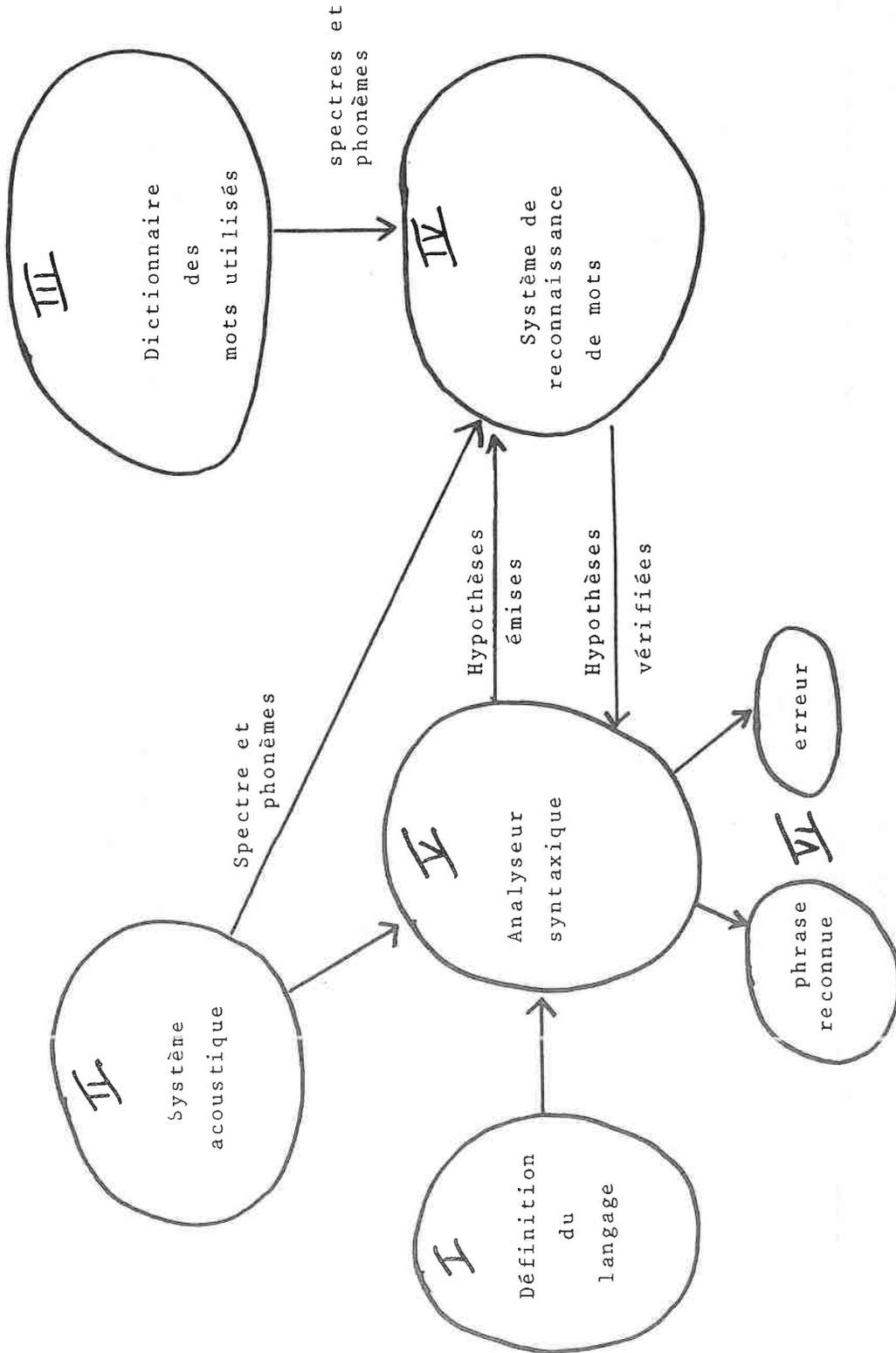


Figure B-1-3-

SYSTEME DE RECONNAISSANCE.



C H A P I T R E I I

DONNEES ET RESULTATS DU SYSTEME

2-1- REPRESENTATION DE LA PHRASE PRONONCEE :

2-1-1- Informations acoustiques :

Le système acoustique peut fournir une représentation du spectre de la phrase. Ce spectre peut être obtenu par un banc de filtres passe-bande dont les fréquences centrales s'échelonnent en général de 100 à 7 000 Hz échantillonnées à une fréquence de 100 Hz. Ce résultat peut être obtenu aussi de façon numérique par une transformée rapide de Fourier (FFT) suivie d'une simulation du banc de filtre.

Le principal inconvénient, lors de l'utilisation du spectre ainsi obtenu, est dû à l'importance de l'information à stocker ; néanmoins il paraît intéressant de s'en servir pour confirmer la chaîne phonémique, spécialement en cas de mauvaise segmentation. Nous avons prévu cette possibilité dans le cadre de notre système, mais nous n'exploitons pas, pour l'instant, ces informations.

D'autres informations fournies par le niveau acoustique doivent aussi pouvoir être prises en considération par le système. Il s'agit essentiellement des indices prosodiques et plus spécialement des variations de la mélodie, que l'on peut détecter par programme ou par un dispositif câblé (47). On peut ainsi obtenir des informations intéressantes, en particulier sur le type de la phrase, interrogative ou affirmative, et envisager l'utilisation de ces indices en liaison avec un module de simplification de la grammaire du langage. Cela devrait permettre d'accélérer l'analyse syntaxique en limitant les diverses possibilités.

2-1-2- La chaîne phonémique :

Pour chaque élément de la chaîne phonétique représentant une phrase à reconnaître, on ne considère pas seulement, un phonème, mais les k phonèmes les plus probables fournis par le niveau acoustique ; ceci augmente la complexité du module de reconnaissance de mots (paragraphe B-4-3), mais aussi les chances de succès, car la probabilité d'avoir le phonème exact est alors bien plus importante. Le nombre de phonèmes pris en compte a été limité à trois, ce qui semble un bon compromis entre la validité de l'information reconnue et la possibilité pour l'algorithme de recherche lexicale de travailler en temps réel. En fait, ces phonèmes sont classés en fonction d'un poids acoustico-linguistique affecté par le niveau de reconnaissance acoustique ; il faut donc plutôt parler de présomptions de présence de phonèmes.

Une chaîne phonémique de LG éléments sera représentée par un tableau GRAPH de dimension $LG \times 3$. On considérera qu'une telle suite se termine toujours par au moins deux positions représentant des silences et, si pour un phonème particulier, le niveau acoustique donne moins de trois possibilités, les positions correspondant aux phonèmes inexistants seront codées par -1.

En résumé, la chaîne phonémique d'entrée est donnée par 2 informations :

LG : longueur de la chaîne.

GRAPH ($LG, 3$) tableau représentant la chaîne.

L'écriture phonétique utilise 34 phonèmes, y compris le silence, et le tableau B-2-1- donne la liste de ces phonèmes avec leur mot clé et la codification utilisée.

TABLEAU B-2-1-

Phonème	mot clé	codification
silence		1
ə	plat	2
ʀ	rue	3
ʁ	lent	4
e	blé	5
s	sous	6
i	il	7
ɛ	lait	8
ə	le, peu	9
ʀ	cou	10
t	tas	11
p	pas	12
d	dans	13
m	ma	14
ɑ̃	an	15
u	nous	16
o	ou	17
v	Vie	18
y	nu	19
ɔ̃	on	20
o	eau	21
ʒ	je	22
ɔ	bol	23
ɛ̃	lin	24
f	feu	25
b	bon	26
w	voir	27
ɣ	huit	28
œ	heure	29
z	zéro	30
j	bailler	31
ʃ	chat	32
g	gare	33
œ̃	un	34

1	v 18	z 30	3 22	13	a 2	à 9	e 5
2	à 9	a 2	e 5	14	m 14	b 26	d 13
3	v 18	3 22	z 30	15	a 2	o 21	à 9
4	u 17	γ 19	o 21	16	b 26	d 13	g 33
5	u 17	à 9	o 21	17	a 2	o 21	à 9
6	b 26	d 13	g 33	18	m 14	â 15	ô 20
7	ε 8	i 7	e 5	19	a 2	o 21	e 5
8	e 5	ε 8	i 7	20	e 4	z 3	-1
9	p 12	t 11	k 10	21	b 26	d 13	g 33
10	a 2	e 5	à 9	22	e 5	ε 8	i 7
11	z 3	e 4	e 5	23	1	1	1
12	e 5	à 9	â 15	24	1	1	1

"Je voudrais parler à Madame ALBERT"

Exemple de donnée en entrée

figure B.2.2.

2-2- DEFINITION DU LANGAGE :2-2-1- Niveau syntaxique : la grammaire.

Le premier niveau de définition du langage est composé d'informations syntaxiques regroupées dans une grammaire qui sera représentée dans le formalisme de Backus-Naur. A titre d'exemple, la figure B-2-5 donne la grammaire du langage composé de l'union des $a^{n+p} b^p$, $a^{n+p} c$, $a^{n+p} b^p d$ (avec $n > 0$ et $p > 0$).

$$\begin{aligned} X &::= \langle A \rangle \langle B \rangle \langle D \rangle / \langle A \rangle \langle B \rangle \\ A &::= a \langle A \rangle / a \\ B &::= a \langle B \rangle b / ab \\ D &::= c / d \end{aligned}$$
figure B-2-5-

Une phrase de codification, pouvant être automatisée, nous permet de représenter cette grammaire de manière directement exploitable par l'analyseur syntaxique.

La codification retenue est (48) :

a) pour les terminaux : des entiers positifs compris entre TI et T ($0 < TI < T$)

A ce niveau, nous distinguons deux types de terminaux, les "terminaux vrais" et les "terminaux de type lexique", qui correspondent à deux représentations différentes dans le dictionnaire.

- les terminaux de type lexique (définition de dictionnaire, d'Annuaire, de liste, etc...) sont ceux provenant de règles du type :

(1) $AG ::= a_1 / a_2 / \dots / a_n$

telles que le membre droit de la règle se compose uniquement d'une suite d'alternatives contenant chacune un seul élément de type terminal. Dans ce cas, le membre gauche sera considéré comme un terminal de type lexique : pour la règle donnée en (1), AG représente un tel terminal. Ce type de terminal correspond assez bien aux unités lexicographiques de type identificateur rencontrées en compilation. Dans la grammaire B-2-5- D est de ce type.

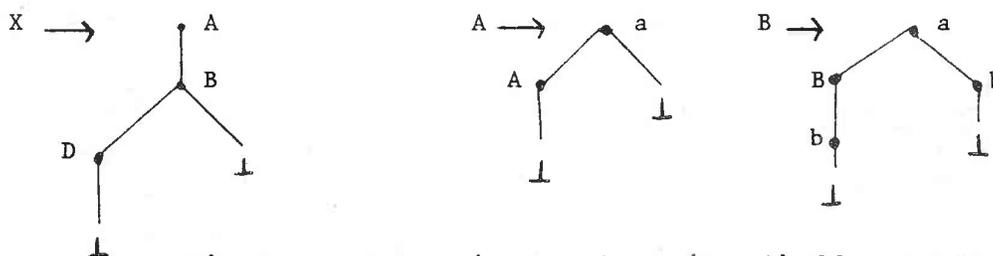
- les terminaux vrais correspondent aux autres terminaux à l'exclusion des terminaux intervenant dans des règles de type (1)).

On codera de TI à TV "les terminaux vrais" de TV + 1 à T "les terminaux de type lexique"

b) pour les non terminaux : des entiers variant de T + 1 à N ($T < N$),

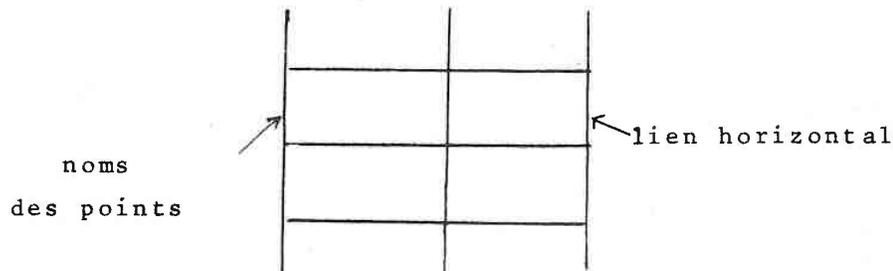
l'Axiome de la grammaire étant noté T + 1

c) pour la grammaire : une ramification par symbole non terminal ; pour la grammaire donnée à la figure B-2-5- on obtient donc :



N'ayant pas besoin par la suite d'affecter de numéros aux différentes règles, on notera '↓' la fin d'une règle.

Ces différentes ramifications sont, à leur tour, représentées par leur matrice d'enchaînement (49), avec lien vertical par continuité ; les éléments n'ayant pas de lien vertical sont les terminateurs '↓' : on les notera par un nombre négatif.



Nous regroupons toutes ces matrices d'enchaînement en une seule. L'absence de lien horizontal est indiqué par un zéro.

La figure B-2-6- donne la représentation de la grammaire présentée à la figure B-2-5-.

indice	nom	lh
1	A	0
2	B	0
3	D	4
4	-1	0
5	a	0
6	A	7

indice	nom	lh
7	-1	0
8	a	0
9	B	10
10	b	0
11	-1	0
12		

Figure B-2-6-

Cette matrice sera représentée par deux tableaux : NOM et LH. Il faut connaître de plus, pour chaque non-terminal A, la racine de la ramification, c'est-à-dire l'indice où commence les règles relatives à A : ces indices seront données dans un tableau DEB (T + 1 : N) : DEB(A) correspond au début des règles relatives à A.

2-2-2- Niveau morphologique : le dictionnaire.

Le dictionnaire regroupe les représentations phonémiques de référence des différents terminaux du langage. On utilise deux types de représentation suivant qu'il s'agit de "terminaux vrais" ou de "terminaux de type lexique".

a) cas des terminaux vrais.

Ils sont représentés par un tableau $TER(LT, 3)$;
 LT représente la longueur du terminal.

L'élément $TER(I, *)$ caractérise le $i^{\text{ème}}$ phonème :

$TER(I, 1)$ donne le phonème souhaité ; si
 $TER(I, 1) < 0$, cela note une possibilité
d'élimination de ce phonème. $TER(LT, 1)$ sera
toujours égal à zéro et servira à indiquer
la fin du mot.

$TER(I, 2)$ donne une autre valeur possible au même
phonème pouvant apparaître par substitution
du phonème $TER(I, 1)$

$TER(I, 3)$ indique, en cas d'insertion possible d'un
phonème avant le phonème I , la valeur de
ce phonème ; sinon $TER(I, 3) = 0$

A titre d'exemple, la figure B-2-7- donne une repré-
sentation lexicale du mot "PETITE" sous forme d'un graphe d'une
part et d'un tableau $TER(LT, 3)$ d'autre part.

L'ensemble des tableaux représentant les différents
terminaux seront fusionnés en un seul ; on a donc besoin d'un
tableau d'entrée $TERMI (RI : TV)$ tel que $TERMI (a)$ donne
l'indice du début de la représentation de a .

b) cas des terminaux de type lexique :

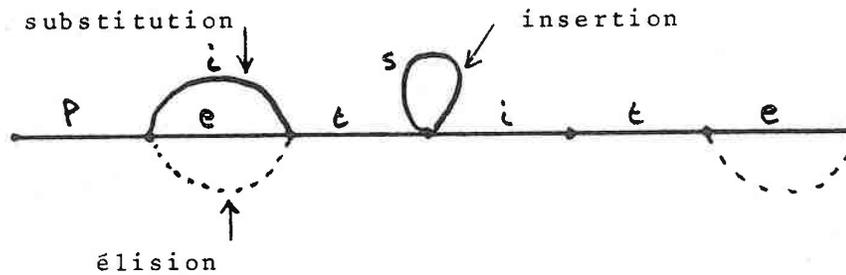
Chaque terminal de type lexique comporte plusieurs
réalisations distinctes qui sont représentées par une informa-
tion arborescente codifiée dans un tableau $PHOLEX (L, 3)$
(L étant la longueur du tableau correspondant à l'arbre des
réalisations).

Pour l'élément I du tableau :

$PHOLEX (I, 1)$ donne la valeur du phonème ;
si $PHOLEX (I, 1) < 0$, cela signifie la
fin de la chaîne phonémique correspondant
à une réalisation du terminal ; dans
ce cas $PHOLEX (I, 1)$ donne l'élimination
possible en fin de mot.

$PHOLEX (I, 2)$ donne la valeur du pointeur indiquant
le phonème voisin de $PHOLEX (I, 1)$ dans

a) graphe représentant le mot



b) tableau correspondant

LT = 7

1	P	12	0	0	
2	[e]	-5	i	7	0
3	t	11	0	0	
4	i	7	0	s	6
5	t	11	0	0	
6	[e]	-5	0	0	
7		0	0	0	

figure B.2.7.

Représentation lexicale du "terminal vrai".

"P E T I T E"

1	d	13	6	2
2	Y	19	0	3
3	P	12	11	4
4	ô	20	0	5
5		0	0	101
6	b	26	15	7
7	u	17	2	8
8	ʃ	32	0	9
9	ε	8	0	10
10		0	0	103
11	t	11	12	4
12	r	3	13	13
13	a	2	0	14
14		0	0	102
15	a	2	20	16
16	e	4	17	17
17	b	26	0	18
18	ε	8	0	19
19	[r]	-3	0	104
20	e	5	0	16

↓ ↓ ↓
 réalisation pointeur pointeur
 du phonème vers vers
 le voisin le suivant

représentation du terminal
de type lexique NOM ayant
quatre réalisations.

Dupont (codé 101)

d Y P ô
b t

Durand (codé 102)

d Y [r] â
b

Bouchet (codé 103)

b u ʃ ε

Albert (codé 104)

a [e] b ε [r]
e

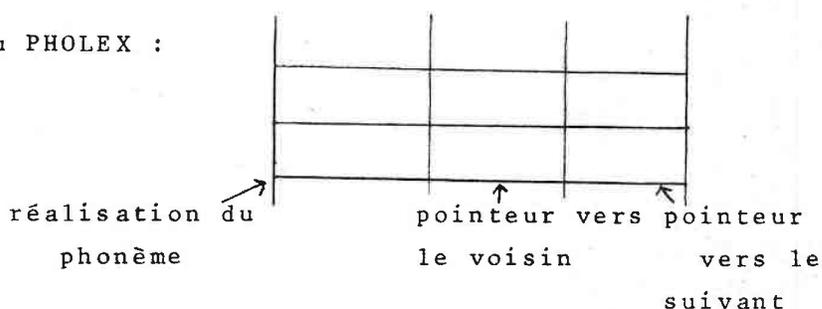
Représentation lexicale d'un terminal de type lexique.

Figure B.2.8.

la ramification. L'absence de voisin est codée par zéro.

PHOLEX (I, 3) donne le pointeur vers le phonème suivant sauf dans le cas où $\text{PHOLEX (I, 1)} \leq 0$; on a alors le numéro de la réalisation du terminal.

tableau PHOLEX :



A titre d'exemple la figure B-2-8- donne la représentation d'un terminal de type lexique possédant 4 réalisations.

Les tableaux correspondant aux divers terminaux de type lexique sont fusionnés en un seul ; on a de plus un tableau LEXIQ (TV + 1, T) tel que LEXIQ (A) donne le début de la représentation du terminal de type lexique A.

2-2-3- Niveau sémantique :

Actuellement, nous disposons de très peu d'outils pour caractériser un langage au niveau sémantique, et c'est là un vaste domaine de recherche en informatique. Les solutions proposées, traitement par attributs ou par double grammaire, sont encore beaucoup trop lourdes pour pouvoir être utilisées dans le cadre de systèmes de communication H-M en temps réel.

Nous avons donc été amenés à reporter l'utilisation de la plupart des informations sémantiques définissant le contexte du langage utilisé au niveau d'une phase de dialogue propre à chaque application. Néanmoins, si nous voulons obtenir

une certaine liberté par rapport à la syntaxe (possibilité d'adjonction ou de suppression de certains mots), il nous est nécessaire de prendre en compte certaines informations sémantiques au cours du traitement syntaxique. Nous nous sommes limités dans le cadre de ce système à définir une liste de mots, caractéristiques sur le plan sémantique, qui est une donnée prise en compte au même titre que la grammaire du langage. Cette liste regroupe les mots sans lesquels, dans le contexte d'une application, on ne peut rien comprendre de l'énoncé oral de départ. Ce sera, par exemple, les noms de l'annuaire pour un standard téléphonique, des mots clés, les noms de fichiers ou d'enregistrements pour une base de données, etc...

2-3- DEFINITIONS DES RESULTATS SOUHAITES :

2-3-1- Différence entre reconnaissance et compréhension :

Le développement des recherches, en vue de la reconnaissance du discours continu a permis de mettre en évidence la distinction entre deux types de systèmes :

a) les systèmes à reconnaître la parole ("speech recognition systems), dans lesquels on essaye de reconstituer élément par élément le message prononcé. Dans ce cas, une erreur de reconnaissance d'un mot entraîne l'abandon du travail pour toute la phrase.

b) les systèmes à comprendre la parole ("speech understanding systems") dans lesquels on s'attache à reconnaître le sens global du message, même si des erreurs subsistent au niveau de la reconnaissance d'un ou plusieurs mots de la phrase. De tels systèmes utilisent souvent des techniques se rapprochant de celles utilisées en intelligence artificielle, spécialement en ce qui concerne l'usage fréquent de retours en arrière ("back-tracking").

Jusqu'ici, les systèmes existant sont surtout du type (a), et l'étude de systèmes de type (b) s'est surtout développée aux Etats-Unis.

Il semble pourtant que la plupart des applications déjà citées (partie A, chapitre 3) correspondent à des systèmes de type (b). On peut donc s'attendre à un développement des recherches dans ce sens au cours des années à venir.

Pour notre part, après avoir réalisé un système de type (a) (50), nous avons abordé le problème de la compréhension, et nous verrons au chapitre 4 comment poursuivre la reconnaissance en cas d'erreur sur un terminal.

2-3-2- Le système de reconnaissance considéré comme un processus provoquant une action.

a) Principe :

Le but final d'un système de reconnaissance ou de compréhension de la parole est en général de faire accomplir par l'ordinateur l'action demandée dans le message reçu : commande d'une machine, branchement d'une ligne téléphonique, traitement sur une base de données... Pour ce faire, le système doit connaître le contexte exact de l'application et être capable de déterminer le sens du message reçu. Il faut donc prévoir, après la reconnaissance proprement dite, une phase d'exploitation des résultats propres à chaque application.

b) Nécessité d'un dialogue :

Mais avant de provoquer l'action correspondant à la demande orale, il faut s'assurer de la validité des résultats fournis par le système de reconnaissance. Pour augmenter la fiabilité de la réponse, il paraît donc intéressant, voire nécessaire, de passer par une phase de dialogue (51). De plus, peut-être peut on trouver dans ce dialogue un moyen d'humaniser le contact avec de tels systèmes automatisés.

Il semble difficile, voire impossible de formaliser ce dialogue dans le cas général, car il est très lié au contexte particulier de chaque réalisation. Nous avons donc pensé mettre en place une procédure propre à chaque application : elle utilise les résultats de la reconnaissance et s'efforce de lever les ambiguïtés et de traiter les erreurs. Nous verrons en partie C un exemple de telles procédures dans le cas particulier de l'automatisation d'un standard téléphonique. Enfin, dans certains cas, il faut prévoir un recours à un opérateur humain, ce que l'on retrouve dans presque tous les systèmes automatisés.

C H A P I T R E I I I



L 'ANALYSEUR SYNTAXIQUE

3-1- RAPPELS AU SUJET DE L'ANALYSE SYNTAXIQUE :

Dans la suite de cette étude, les notions de grammaire, langage engendré par une grammaire, ramifications et analyse syntaxique sont supposées connues. Seul un souci de clarification de l'exposé nous amènera à faire quelques rappels ; pour le reste on pourra se référer aux travaux de C. PAIR (48), (49), (52) ainsi qu'aux ouvrages de AHO et ULLMAN (68).

3-1-1- Rappels de notions utilisées dans la suite :

Une grammaire à contexte libre G est formée d'un quadruplet $(T, N, ::=, X)$ où :

- T est le vocabulaire terminal (ensemble des symboles terminaux ou terminaux) ;
 - N est le vocabulaire non terminal (ensemble des symboles non terminaux ou non terminaux) ;
 - T et N sont des ensembles disjoints ; on notera $V = T \cup N$;
 - X est l'axiome de la grammaire ($X \in N$) ;
 - $::=$ est une relation binaire appelée relation de production entre N et V^* (ensemble des mots sur V)
- Un couple (A, α) tel que $A ::= \alpha$, $A \in N$ et $\alpha \in V^*$ s'appelle une règle de la grammaire.

Il est possible de définir entre les éléments de V et V^* un certain nombre de relations qui seront utiles dans la suite.

a) relation de dérivation :

La relation se réécrit " \rightarrow ", est définie dans V^* par:

$$\alpha \rightarrow \beta \iff (\exists A \in N, \exists \lambda \in V^*, \exists \lambda' \in V^*, \exists \mu \in V^*)$$

$$(\alpha = \lambda A \lambda', \beta = \lambda \mu \lambda' \text{ et } A ::= \mu)$$

la fermeture transitive de \succ est notée \succ^*

lorsque $\alpha \succ^* \beta$ on dit que β dérive de α

Le langage engendré par une grammaire G est l'ensemble des mots de T^* qui dérivent de l'axiome X .

b) relation "initiale" :

Soit \mathcal{J} la relation dans V définie par

$$\alpha \mathcal{J} \beta \iff (\exists \varphi \in V^*, \exists \psi \in V^*) (\alpha ::= \varphi \beta \psi \text{ et } \varphi \succ^* \Lambda)$$

Λ représente la chaîne vide

la fermeture transitive de la relation \mathcal{J} est notée \mathcal{J}^+

la fermeture réflexive et transitive de la relation \mathcal{J} est notée \mathcal{J}^*

Lorsque $A \mathcal{J}^+ B$, on dit que B est initial stricte de A

Lorsque $A \mathcal{J}^* B$, on dit que B est initial de A

S'il existe un symbole qui est initial stricte de lui-même, la grammaire est recursive à gauche.

c) la relation finale :

Soit la relation \mathcal{V} définie par :

$$A \mathcal{V} B \iff (\exists \varphi \in V^*, \exists \psi \in V^*) (A ::= \varphi B \psi \text{ et } \psi \succ^* \Lambda)$$

Comme pour la relation \mathcal{J} , il est possible de définir la fermeture transitive de \mathcal{V} : \mathcal{V}^*

d) la relation successeur :

Soit \mathcal{P} la relation définie dans V par :

$$A \mathcal{P} B \iff (\exists C \in N, \exists \varphi \in V^*, \exists \mu \in V^*, \exists \psi \in V^*) \\ (C ::= \varphi A \mu B \psi \text{ et } \mu \succ^* \Lambda)$$

e) la relation voisin :

Soit \mathcal{V}_s la relation dans V définie par $\mathcal{V}_s = \tilde{\mathcal{P}}^{-1} \mathcal{P} \mathcal{J}^*$

$\tilde{\mathcal{P}}^{-1}$ est la relation inverse de $\tilde{\mathcal{P}}$; $\tilde{\mathcal{P}}^{-1}$ en est la fermeture réflexive et transitive

$$(A \mathcal{V}_s B) \iff (\exists C \in V, \exists D \in V) (C \mathcal{P} D \text{ et } C \tilde{\mathcal{P}}^* A \text{ et } D \mathcal{J}^* B)$$

Alors :

$a \mathcal{V}_s b$ et $a \in T$ et $b \in T \iff$ les terminaux a et b peuvent se suivre dans un mot du langage.

3-1-2- Définition de l'analyse syntaxique :

Etant donné une grammaire G , d'alphabet terminal T et une phrase α , effectuer l'analyse syntaxique de α pour la grammaire G , c'est trouver les ramifications de r engendrées par G telles que α soit le mot des feuilles.

Si α n'appartient pas au langage L engendré par G , on ne trouve aucune ramification r ; si α appartient à L on en trouve une ou plusieurs (l'existence de plusieurs ramifications signifie que la phrase α est ambiguë pour la grammaire G). L'analyse syntaxique résoud donc en particulier le problème de la reconnaissance des mots de L .

Parmi les méthodes d'analyse syntaxique travaillant de la gauche vers la droite, on distingue deux grandes méthodes :

- l'analyse descendante (top-down) qui construit la ramification à partir de l'axiome.

- l'analyse ascendante (bottom-up) qui construit la ramification à partir du mot α pour remonter vers l'axiome.

3-1-3- Définition de l'analyse syntaxique descendante :

L'Analyse syntaxique descendante consiste à appliquer un algorithme qui, partant de l'Axiome, construit de proche en proche la ramification admettant comme mot engendré la donnée α .

Le résultat souhaité correspond à l'écriture de la phrase ainsi traitée; pour cela nous utiliserons une procédure sortir (a) qui imprime le terminal traité à la suite de la partie gauche de la phrase préalablement imprimée.

Cet Algorithme consiste essentiellement en une procédure analyse A , pour chaque $A \in V$, qui lit une partie β de la donnée et imprime le mot β généré par les ramifications de racine A . Le résultat est immédiat pour $A \in T$: c'est A si la donnée à lire commence par A . Pour $A \in N$, la procédure doit choisir la règle principale de la ramification parmi les règles de

premier membre A : elle est conduite à appeler les procédures d'analyse associées aux symboles du second membre. C'est donc une procédure indéterministe et en général récursive. On peut décrire ainsi analyse A :

- pour un terminal $A \in T$:

```

début   lire(c) ;
           si   = A alors  sortir(A)
           sinon erreur
fin

```

- Pour un non terminal A :

```

début choix (d'une trègle  $A_K ::= B_{K1} \dots B_{Kq_K}$  telle que
            $A_K = A$ );
           analyse  $B_{K1}$  ; Analyse  $B_{K2}$ ; ...; Analyse  $B_{Kq_K}$  ;
fin

```

Remarque : "c" désigne le caractère suivant à lire dans la donnée. On entend ici par caractère une unité lexicographique.

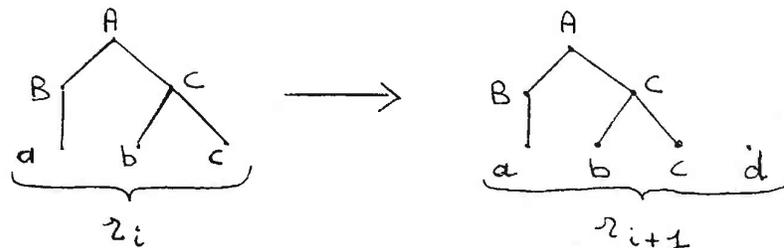
3-1-4- Analyse syntaxique ascendante :

L'analyse syntaxique ascendante construit les ramifications cherchées "de bas en haut et de la gauche vers la droite". La ramification finale r sera obetnue comme dernier élément d'une suite de ramifications $r_0=1, r_1, r_2, \dots, r_n=r$ qui en sont des parties "en bas à gauche".

Pour passer de la ramification r_i à r_{i+1} , on a le choix entre deux possibilités :

(1) soit ajouter une feuille de r à droite de r_i

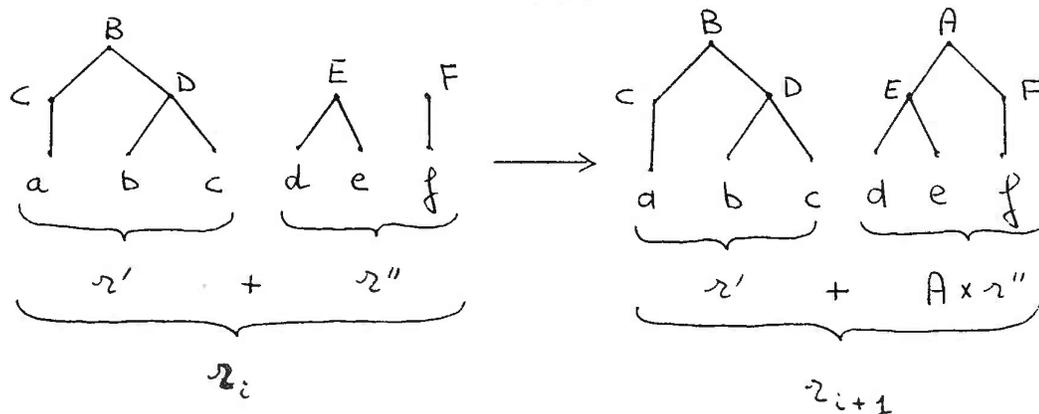
$$\text{ex : } r_{i+1} = r_i + d \quad (*)$$



(2) soit faire une "réduction à droite"

$$r_i = r' + r''$$

$$r_{i+1} = r' + A * r'' \quad (*)$$



Cela correspond aussi à un algorithme indéterministe, car il y a choix à chaque étape entre (1) et (2) et dans (2) choix de la ramification r'' sur laquelle porte l'opération d'enracinement.

(*) on utilise les notations de (48) : '+' désigne la concaténation de ramification et '*' l'enracinement.

3-1-5- Solution retenue :

Le système que nous avons mis au point utilise une analyse descendante : nous avons voulu dans un premier temps valider cette approche qui s'appuie plus sur le niveau syntaxique et ne considère la phase de reconnaissance phonémique d'un terminal que comme vérification du chemin pris lors de l'analyse syntaxique. Un système basé sur l'analyse ascendante aurait une démarche inverse : le centre en serait le module de reconnaissance phonémique du terminal et l'analyse servirait à infirmer ou confirmer les résultats obtenus. Une poursuite intéressante de ce travail consisterait à réaliser la deuxième solution et comparer les résultats obtenus suivant les deux approches en fonction de critères de fiabilité et de rapidité.

D'autre part, nous avons posé deux restrictions sur la grammaire du langage à reconnaître :

a) nous considérerons que la grammaire n'est pas récursive à gauche, c'est-à-dire qu'il n'existe aucun symbole non terminal, initiale stricte de lui-même, soit :

$$(\nexists A \in N, \nexists \varphi \in V^* / A \xrightarrow{*} A\varphi)$$

Cette condition est nécessaire et suffisante pour que l'algorithme d'analyse syntaxique soit borné ; de plus elle n'est en aucun cas restrictive, car on démontre que, pour toute grammaire, on peut trouver une grammaire non récursive à gauche qui lui soit équivalente.

b) d'autre part, afin de faciliter la mise en place de l'algorithme, nous supposerons qu'il n'y a pas de règle vide ; c'est-à-dire : $(\nexists A \in N / A ::= \Lambda)$

Il en résulte que : $(\nexists A \in N / A \xrightarrow{*} \Lambda)$

3-2- PRINCIPALES DIFFERENCES PAR RAPPORT AUX PROBLEMES DE COMPILATION.

L'utilisation de l'analyse syntaxique dans des applications telles que la compilation est maintenant un problème bien connu en informatique (48). C'est pourquoi il semble intéressant de noter les différences fondamentales qui apparaissent dans les problèmes de reconnaissance de formes et plus spécialement de reconnaissance de la parole au niveau syntaxique, afin de mieux mettre en évidence les outils spécifiques qu'il faut alors utiliser.

Les trois différences fondamentales concernent :

- (i) la définition des terminaux,
- (ii) l'inadaptation des algorithmes avec lecture à l'avance de k caractères pour lever l'indéterminisme,
- (iii) la possibilité d'obtenir plusieurs résultats distincts, même dans le cas d'une grammaire non ambiguë.

3-2-1- Définition des terminaux :

Une des grandes caractéristiques de ce type de problème concerne la définition des terminaux. Alors qu'en compilation on arrive toujours à détecter avec certitude la présence ou l'absence d'un terminal, il nous faut ici travailler avec des présomptions de présence. En effet, comme nous l'avons vu au chapitre 2, la chaîne phonémique représentant la phrase à reconnaître est formée à chaque instant des trois phonèmes les plus probables et peut, de plus, être assez fortement entachée d'erreurs.

Même en cas de chaînes phonémiques non entachées d'erreurs, le problème des ambiguïtés possibles sur les terminaux reste posé. Celles-ci peuvent en effet résulter soit de mots possédant des homonymes (sain/saint/sein) soit de phrases phonétiques offrant la possibilité de segmentations multiples en mots ("Gal, amant de la reine, alla, tour magnanime,

galamment de l'arène à la tour Magne, à Nîmes", Victor Hugo). Dans sa grammaire française, CRESSOT appelle "paraphonie" de telles figures de langage.

La lecture d'un terminal ou l'analyse lexicographique n'est donc pas un problème simple et le résultat fourni ne peut en aucun cas être considéré comme absolument sûr. Ceci nécessite une adaptation de la procédure d'analyse syntaxique afin de tenir compte le mieux possible des ambiguïtés existant au niveau des terminaux : il faut en particulier prévoir l'éventualité de retour arrière dans le cas de mauvaise interprétation d'une unité morphologique.

3-2-2- Inadaptation des algorithmes permettant de supprimer l'indéterminisme :

Nous avons défini au paragraphe 3-1- la procédure d'analyse syntaxique de manière indéterministe (introduction d'instruction de type choix). En fait, l'indéterminisme est levé le plus souvent grâce à des algorithmes utilisant une lecture à l'avance de k caractères ou terminaux (48) : on obtient des systèmes acceptant des grammaires $LL(K)$. De telles méthodes permettent de lever l'indéterminisme sur la grammaire lors de l'analyse, mais pour nous il y a de plus indéterminisme sur les terminaux et dans ce cas, la lecture à l'avance de K caractères ne permet pas de le supprimer.

Une autre manière de lever l'indéterminisme réside dans le fait d'utiliser une pile, et de parcourir les différents chemins possibles en effectuant les retours en arrière nécessaires en cas d'échec de l'analyse. Cette méthode risquait d'augmenter considérablement le temps de calcul, et, pour cette raison nous l'avons rejetée comme incompatible avec notre objectif qui était d'obtenir un système aussi proche que possible du temps réel.

Nous avons donc été amenés à modifier les algorithmes classiques de manière à traiter le plus grand nombre de langages sans pour autant trop alourdir le temps de calcul. L'algorithme

mis en place est présenté au cours de l'étude de l'Analyseur syntaxique réalisé (paragraphe 3-3-) ; il permet de traiter tous les langages à contexte libre.

3-2-3- Fossibilités d'obtenir plusieurs résultats distincts :

La dernière différence entre les problèmes de compilation et ceux de reconnaissance de la parole concerne enfin les résultats fournis par l'analyse syntaxique. Alors que dans le premier cas on obtient un résultat unique, si du moins la grammaire n'est pas ambiguë, il nous faut envisager ici la possibilité supplémentaire d'obtenir plusieurs résultats distincts correspondant aux différents cas de mauvais découpages ou de confusions de mots. Pour ce faire, il faudrait traiter toutes les possibilités au niveau de la grammaire, ce qui, comme nous l'avons déjà noté, alourdit considérablement l'algorithme. En fait, nous nous limiterons à la solution la plus probable au vu des scores de reconnaissance des différents terminaux, et nous considérerons comme résultat d'une part la phrase reconnue, d'autre part la pile de gestion des retours-arrière qui indique les points de reprise des autres possibilités et les ambiguïtés possibles. Une phase de dialogue reprendra ces résultats pour les valider avant de déclencher l'action correspondante, but final d'un tel système.

3-3- MODULE D'ANALYSE SYNTAXIQUE REALISE :

Lors de la réalisation de l'analyseur syntaxique, le problème consiste essentiellement à trouver une procédure équivalente à la procédure indéterministe et récursive présentée au paragraphe 3-1-3- permettant, le traitement d'une phrase dans un temps minimum (pour obtenir des résultats proches du temps réel) et une programmation dans le langage choisi : FORTRAN. Le premier objectif nécessite de réduire au maximum l'indéterminisme sans effectuer le traitement correspondant à toutes les possibilités. Quant au second objectif, il nécessite la transformation de la procédure en une procédure déterministe et non récursive.

3-3-1- Réduction de l'indéterminisme :

Comme nous l'avons déjà vu, il existe dans le cadre de la reconnaissance automatique deux causes d'indéterminisme, l'une liée à la grammaire, l'autre liée à la définition de terminaux. Nous ne pouvons pas supprimer cet indéterminisme, mais nous avons essayé de le réduire au maximum.

L'idée directrice de l'algorithme utilisé consiste à ne traiter l'indéterminisme qu'à deux niveaux :

- par l'analyseur lexicographique
- par un calcul de score de reconnaissance.

a) réduction de l'indéterminisme par l'analyseur lexicographique :

Il s'agit en fait de traiter le choix entre différentes règles de la grammaire par un choix entre différents terminaux possibles, effectué par l'analyseur lexicographique. Un tel traitement est possible s'il n'existe pas plusieurs choix différents qui utilisent le même terminal, ce qui, exprimé de façon différente, signifie qu'à un moment donné la prise en compte d'un terminal permet de caractériser de façon non ambiguë le contexte droit à analyser.

C'est le cas en particulier lors du traitement des langages LL(1) : la lecture à l'avance d'un terminal permet de choisir de façon non ambiguë la règle à traiter.

Exemple : $A ::= \langle B_1 \rangle \dots\dots\dots B_n \quad (1)$

$A ::= \langle C_1 \rangle \dots\dots\dots C_n \quad (2)$

Si le langage est de type LL(1), l'ensemble des initiales de B_1 et l'ensemble des initiales de C_1 sont disjoints, et le choix entre (1) et (2) se fait en fonction du prochain terminal à traiter.

Le système a été généralisé pour permettre d'accepter une classe de langage plus importante que les langages générés par des grammaires LL(1). Nous avons été amenés à définir un analyseur acceptant les langages définis par des grammaires de GREIBACH (53) qui sont elles aussi compatibles avec la condition de choix énoncée précédemment.

Rappel sur les grammaires réduites de GREIBACH

Considérons la grammaire

$$X ::= aX / bYX / aZY$$

$$Y ::= cY / c$$

$$Z ::= aZb / ab / d$$

Cette grammaire définit les langages $L(X), L(Y), L(Z)$, solution du système d'équations associées.

Pour savoir si un mot x appartient à $L(X)$, il suffira, si $x=a$, de vérifier que $\alpha \in L(X)$ ou $\alpha \in L(Z) \cap L(Y)$ et si $x=b$ que $\alpha \in L(Y)L(Z)$. La connaissance de $x \in T$ caractérise donc de manière non ambiguë le contexte droit à analyser : cela est vrai pour toutes les grammaires où les second membres de règle sont tous de la forme

$$L\alpha \text{ avec } L \in T \text{ et } \alpha \in (TUN)^*$$

De telles grammaires sont dites sous la forme réduite de GREIBACH, du nom de celle qui, la première, démontra le théorème suivant :

"Toutes les grammaires définissant un langage de CHOMSKY admettent une grammaire équivalente sous la forme réduite de GREIBACH".

Ainsi, dans le cas de langages définis

soit par une grammaire LL(1) (et par extension dans le cas de langages LL(1)),

soit par une grammaire réduite de GREIBACH,

l'instruction CHOIX de la procédure indéterministe peut se décomposer en deux étapes :

(i) détermination de tous les couples (a, k) tels que

$$(\alpha \in T, A_k = B_{k1} \dots B_{kq_k}, A = A_k \text{ et } B_{k1} \gamma^* \alpha)$$

(ii) choix d'un couple (a, i) par l'analyseur lexicographique en fonction des scores de reconnaissance des différents terminaux intervenant dans les couples définis en (a).

On peut donc réécrire la procédure indéterministe analyse A décrite en 3-1-3- de la façon suivante :

pour $A \in N$

début

déterminer l'ensemble des couples (a, k) tels que :

$$(\alpha \in T, A_k = B_{k1} \dots B_{kq_k}, A = A_k, B_{k1} \gamma^* \alpha);$$

choix (de c, i) pour l'analyseur lexicographique) ;

CHOIX = VRAI ;

analyse B_{k1} ; analyse B_{k2} ; ; analyse B_{kq_k} ;

fin

pour $A \in T$

début

si CHOIX=.FAUX. alors choix (de c par l'analyseur lexicographique)

si C=A alors sortir (A)

sinon erreur

fin

b) réduction de l'indéterminisme par calcul de probabilité ou score de réussite :

Dans le cadre de la reconnaissance de la parole, il semble souhaitable d'adjoindre à la phrase reconnue une probabilité de reconnaissance afin de pouvoir abandonner certains chemins au cours de l'analyse syntaxique lorsque cette probabilité devient trop faible ; on peut ainsi réduire aussi l'indéterminisme. Nous nous trouvons dans le cas d'un système qui évolue au cours du temps pour prendre successivement les états E_0, E_1, \dots, E_n correspondant aux moments où sont reconnus 0, 1, ..., n terminaux du langage. Pour cela il faut déterminer

$$P(E_j^n / E_{i_0}^0 \quad E_{i_1}^1 \dots \quad E_{i_n}^n)$$

Ceci conduit en probabilité à traiter une chaîne de Markov qui, dans le cadre de notre application, serait non homogène dans le temps.

Il nous est apparu que les données nécessaires à un tel traitement (ensemble des probabilités initiales, ensemble des probabilités de transition) et le calcul conduisant à la probabilité de la trajectoire utilisée sont beaucoup trop lourds pour l'utilisation que l'on veut faire de ce résultat. Il s'agit pour nous de tester la qualité du contexte gauche déjà traité pour savoir s'il y a lieu ou non de poursuivre le chemin pris jusqu'alors. Nous avons donc décidé d'abandonner ce calcul de probabilité au profit d'un score de réussite (variant de 0 à 1), déterminé de manière empirique.

Soit SC_{i-1} le score cumulé du contexte gauche déjà reconnu, S_i celui du dernier terminal fourni par l'analyseur lexicographique, et SEUIL 3 le seuil à partir duquel nous considérerons que la reconnaissance du dernier terminal infirme ou confirme le contexte gauche déjà reconnu, nous déterminerons le nouveau score de réussite SC_i par le calcul récurrent suivant :

$$SC_0 = 1$$

$$SC_i = \text{MIN}(1, SC_{i-1} + S_i - \text{SEUIL } 3)$$

Si le seuil à partir duquel on décide d'abandonner la reconnaissance est SEUIL 2, le test mis en oeuvre pour réduire l'indéterminisme sera :

si $SC_i < \text{SEUIL } 2$ alors abandon du chemin
sinon poursuite

Nous verrons dans le cadre de l'Analyseur lexicographique comment obtenir S_i . Les valeurs de SEUIL 2 et SEUIL 3 optimales pour obtenir de bons résultats ont été fixées par expérience à 0.5 et 0.8. Elles restent ajustables, car rien ne prouve que ce soit là réellement les meilleures valeurs. De plus, suivant l'application, on peut être amené soit à augmenter ces seuils pour obtenir une plus grande fiabilité de reconnaissance, au risque de voir s'accroître le nombre de phrases rejetées, soit à les diminuer si l'on recherche simplement la phrase la plus proche phonétiquement et syntaxiquement de l'énoncé oral de départ.

3-3-2- Suppression de l'indéterminisme et de la récursivité :

a) traitement de l'indéterminisme :

Comme nous l'avons vu dans le paragraphe précédant, l'algorithme n'est plus indéterministe qu'au niveau de l'analyseur lexicographique. Pour supprimer cet indéterminisme, il suffit d'introduire une pile dans laquelle sont rangés à chaque pas les divers couples validés par l'analyseur

lexicographique. Ce rangement se fera suivant l'ordre croissant des scores de reconnaissance affectés aux divers terminaux reconnus.

Le but de l'analyseur lexicographique est donc de tester chaque couple (a,k) possible et, en fonction des scores de reconnaissance (si $S_i > \text{SEUIL } 2$) de ranger les résultats obtenus dans la pile des hypothèses validées : le résultat pris en compte pour la poursuite de l'analyse syntaxique correspond alors au sommet de pile. Si la reconnaissance ne conduit à la confirmation d'aucun couple (a,k), le résultat (sommet de pile) correspond alors à une hypothèse précédemment placée dans la pile et sa prise en compte provoque un retour en arrière dans l'analyse syntaxique ; il faut donc adjoindre dans la pile le contexte d'analyse syntaxique aux hypothèses validées. Dans le cas où cette pile devient vide, cela indique l'impossibilité de nouveaux retours en arrière et provoque une erreur pour la reconnaissance de la phrase traitée.

b) traitement de la récursivité :

Le traitement de la récursivité se fait lui aussi par la méthode classique d'utilisation d'une pile. Deux cas sont à considérer lors de la suppression de la récursivité :

- (i) l'appel de la procédure : il est remplacé par
 - la sauvegarde sur la pile du contexte de l'analyse en cours,
 - la sauvegarde en sommet de pile du point de retour
 - un branchement au début de la procédure d'analyse, après mise à jour des paramètres correspondant à cet appel.
- (ii) la fin de la procédure : elle est remplacée par :
 - la restauration du point de reprise sauvegardé en sommet de pile lors de l'appel,
 - la restauration à partir de la pile du contexte existant avant l'appel de la procédure,
 - un branchement au point de reprise.

Il est possible d'écrire un programme déterministe et non récursif décrivant l'algorithme que nous venons de présenter ; il utilise deux piles :

- la pile des couples (a,k) sélectionnés ou pile des hypothèses validées, qui sera gérée par l'analyseur lexicographique,
- la pile d'analyse syntaxique utilisée lors de la suppression de la récursivité.

3-4- DESCRIPTION FONCTIONNELLE DE L'ANALYSEUR SYNTAXIQUE :

Dans ce paragraphe, nous nous référerons à la représentation des données présentées en partie B paragraphe 2-2-.

On notera $TP1=T+1$

3-4-1- Création de la matrice IN de la relation \mathcal{J}^*

Au cours de l'Analyse syntaxique nous avons besoin de connaître la relation \mathcal{J}^* ; celle-ci sera créée dès que la grammaire du langage sera connue, et elle sera représentée par une matrice booléenne $IN(T1:N, TP1:N)$ telle que :

si $B \mathcal{J}^* a$ alors $IN(a,B)=.VRAI.$

sinon $IN(a,B)=.FAUX.$

a) création de la matrice IN de la relation \mathcal{J}

Mise à .FAUX. de tous les éléments de IN

Pour I de TP1 à N faire

début $K=DEB(I)$

tant que $K \neq 0$ faire

début $IN(NOM(K), I)=.VRAI.$

$K=LH(K)$

fin

fin

b) construction de la fermeture transitive \mathcal{J}^*

La fermeture transitive de la relation \mathcal{J} s'élabore à l'aide de l'algorithme de WARSHALL (55), (54).

Dans le cadre de notre application on obtient :

```

Pour I de TP1 à N faire
  début Pour J de TP1 à N faire
    début si IN(I,J) alors début
      Pour K de TI à N faire
        IN(K,J)=IN(K,J) UIN(K,I)
      fin
    fin
  fin
  IN(I,I)=.VRAI.
Fin

```

3-4-2- Procédure d'Analyse syntaxique :

La procédure d'Analyse syntaxique a donc pour but d'émettre des Hypothèses sur les mots à reconnaître en fonction du contexte gauche déjà traité. L'algorithme utilisé peut se décrire ainsi (** indiquera le début d'une ligne de commentaire) :

```

** variables utilisées :
**   grammaire du langage tableau NOM, LH, DEB
**   T limite dans la codification entre terminaux et non terminaux
**   J pointe dans la grammaire vers l'élément à
**       analyser
**   IR donne le nombre d'éléments du contexte gauche
**       déjà reconnu
**   NTER : numéro du terminal correspondant à
**       l'hypothèse sélectionnée pour la poursuite
**       de l'analyse syntaxique
**   IN(TI:N,TP1:N) tableau booléen correspondant
**       à la relation  $\gamma^*$ 
**
** INITIALISATION
**   J = DEB(TP1) ; IR=0
** détermination du terminal suivant
**   HYPOTHESE : sous-programme de création des
**       hypothèses
**       il crée NB hypothèses et les range
**       dans un tableau TAB(N,3)
**       indiquant pour chaque hypothèse
**       le nom du terminal, et son contexte d'analyse

```

```

**          indiqué par J et IR.
** RECH : correspond au module d'analyse lexicographique
**        décrit au chapitre 4.
**        il valide les hypothèses rangées dans TAB
**        et en sélectionne une, composée d'un
**        terminal NTER et de son contexte J et IR.

HYP : HYPOTHESE (TAB, NB, J, IR)
RET : RECH(TAB, NB, NTER, J, IR)

** TEST : y-a-t-il une hypothèse validée ?
        si NTER = 0 alors ERR(1)

** Analyse syntaxique
RECØ : B = NOM(J)
        si B > 0 alors début 1
            si B > t alors début 2

** cas où B représente un nom terminal
        si IN(NTER,B) alors début 3
            EMPILER(J)
            J=DEB(B)
            aller à RECØ
                fin 3
            fin 2
            sinon début 4

** cas où B représente un terminal
        si B = NTER alors début 5
            IR = IR + 1
            SORTIR(B,IR)
            J=J+1
            aller à HYP
                fin 5
            fin 4

** traitement des autres possibilités indiquées par le
   lien horizontal

```

```

    si LH(J)≠0 alors début 6
        J=LH(J)
        aller à RECØ
        fin 6
    sinon ERR(2)
    fin 1
sinon début 7

```

**** cas où B correspond à une fin de règle**

```

DEPILER (J)
si J > 0 alors aller à RECØ
    sinon
        si FIN de PHRASE alors ECRIRE (IR)
            sinon aller à RET
    fin 7

```

fin du module d'analyse syntaxique

3-4-3- Procédure d'émission des Hypothèses :

La procédure d'émission des Hypothèses a comme paramètres J et IR qui caractérisent le contexte droit à reconnaître.

Elle fournit comme résultat NB hypothèses rangées dans un tableau TAB. Chaque élément est formé d'un triplet (NTER, J, IR) où NTER donne le numéro du terminal, J et IR caractérisant son contexte.

**** INITIALISATION**

NB = 0

**** Peut-on valider des hypothèses ?**

si Non Fin de phrase alors début 1

RECØ : B = NOM(J)

si B > 0 alors début 2

si B > t alors début 3

**** cas correspondant à un non terminal. On recherche ses**

**** initiaux terminaux que l'on range dans le tableau A.**

INIT(B,A,n)

**** adjonction de ces initiaux comme hypothèses dans TAB**

ADJT(A,n,J,IR,NB)

fin 3

sinon début 4

** cas d'un terminal ; on l'ajoute comme hypothèse dans TAB

A(1) = B

ADJT(A,1,J,IR,NB)

fin 4

** traitement des autres possibilités offertes par la

** grammaire

si LH(J) \neq 0 alors début 5

J = LH(J)

aller à RECO

fin 5

fin 2

sinon début 6

** cas où B correspond à une fin de règle

DEPILER (J)

si J > 0 alors aller à RECO

fin 6

fin 1

fin du module d'émission des hypothèses.

3-4-4- Procédures de service utilisées :

a) gestion des retours-arrière possibles lors de l'Analyse syntaxique.

EMPILER(J): sauvegarde le contexte correspondant à J.

DEPILER(J): restaure le contexte correspondant à J.

b) traitement des erreurs :

ERR(N) : arrête l'analyse syntaxique et la reconnaissance de la phrase, et provoque la mise en route de la phase de dialogue. N indique le type d'erreur ayant provoqué l'appel à cette procédure.

c) gestion de la phrase reconnue :

SORTIR(B,IR) : indique dans un tableau que B a été reconnu en position IR

ECRIRE(IR) : procédure de récupération des résultats de la reconnaissance. Elle correspond dans notre système à un renvoi à la phase de dialogue.

d) gestion des hypothèses :

INIT(B,A,n) : recherche les initiaux de type terminaux de B et les range dans un tableau A ; n indique le nombre de terminaux ainsi sélectionnés

ADJT(A,n,J,IR,NB) : ajoute au tableau TAB n hypothèses dont les numéros de terminaux sont dans A et le contexte fourni par J,IR et la pile des retours arrière.

CHAPITRE IV

L'ANALYSEUR LEXICOGRAPHIQUE (ou MORPHOLOGIQUE)

4-1- BUT DE L'ANALYSEUR LEXICOGRAPHIQUE :

L'analyseur lexicographique a pour but de valider les hypothèses émises par l'analyseur syntaxique et de sélectionner la meilleure pour permettre la poursuite du travail de ce dernier.

On peut y distinguer trois étapes :

a) Il tente la reconnaissance de tous les terminaux émis comme hypothèses et rejete ceux qui provoquent un échec. Ceci se fait grâce à un module de reconnaissance de mot qui compare le début de la sous-chaîne phonémique non encore traitée et les représentations lexicales des différents terminaux. Nous verrons au chapitre V comment est réalisée cette recherche lexicale.

b) Il classe les terminaux reconnus, accompagnés de leur contexte, en fonction du taux de réussite S_I fourni pour chacun par le module de reconnaissance de mots.

c) Enfin, il range ces résultats, triés par ordre croissant des taux S_i , sur la pile des hypothèses valides, et sélectionne le sommet de pile comme point de reprise de l'analyseur syntaxique.

4-1-1- Données en entrées :

Cet analyseur accepte en entrée deux types de données :

- l'ensemble des hypothèses composées chacune d'un couple (numéro de terminal, contexte de l'analyse syntaxique).

Un même terminal peut apparaître plusieurs fois au sein de ces hypothèses. Aussi, pour optimiser le nombre d'appels au module de reconnaissance, il effectue un tri de ces couples en fonction du numéro de terminal comme traitement préliminaire.

- la représentation phonémique de la phrase à reconnaître ainsi qu'un pointeur I indiquant le rang du dernier phonème pris en compte lors de la reconnaissance du contexte gauche des hypothèses à tester.

4-1-2- Résultat_fourni :

Le résultat fourni par l'analyseur lexicographique correspond en général à un point de reprise pour le niveau syntaxique composé d'un numéro de terminal reconnu et de son contexte. En fait, différents cas peuvent se produire, correspondant au niveau syntaxique, à des situations différentes.

La situation la plus favorable est celle où l'une au moins des dernières hypothèses émises a été validée par le module de reconnaissance de mot. La prise en compte du sommet de pile correspond alors à la poursuite normale de l'analyse syntaxique et à une progression dans la phrase à reconnaître.

Dans le cas contraire, deux cas peuvent se présenter suivant que la pile des hypothèses précédemment validées est épuisée ou non. Si cette pile n'est pas vide, on peut déterminer un point de reprise. Il correspond alors à un retour en arrière au niveau de l'analyse syntaxique et de la reconnaissance de la phrase : le système a exploré un mauvais chemin, il lui faut maintenant prendre en compte une autre possibilité qu'il avait sauvegardée lors d'une étape précédente. Il s'agit là d'une stratégie classique en intelligence artificielle.

Enfin, dans le cas où toutes les hypothèses validées ont été utilisées, l'analyseur syntaxique indique simplement qu'aucun terminal n'est reconnu. Tous les chemins ont conduit à l'échec de la reconnaissance : la phrase de départ ne fait donc sans doute pas partie du langage à reconnaître ou bien le traitement au niveau acoustique a été erroné.

4-2- GENERALISATION DE L'ANALYSEUR : LIBERTE PAR RAPPORT A LA GRAMMAIRE.

Lors de la vérification des hypothèses, l'analyseur lexicographique décide s'il faut ou non poursuivre la reconnaissance de la phrase, compte tenu du contexte gauche déjà reconnu. Dans ce but, le module de reconnaissance de mot recherche si la partie de la chaîne phonémique non encore explorée commence par la représentation d'un des terminaux émis comme hypothèses. Le succès d'une telle opération nécessite, en donnée, une phrase syntaxiquement correcte et bien prononcée. Ce n'est, hélas, pas toujours le cas lors d'une conversation orale, à plus forte raison si celle-ci a lieu dans un langage spécialisé à syntaxe précise. Nous avons donc décidé de généraliser ce vérificateur d'hypothèses afin d'acquérir une certaine liberté par rapport à la grammaire et d'obtenir ainsi un système de compréhension de la parole plutôt qu'un système de reconnaissance stricte (cf partie B, § 2-3).

4-2-1- Exposé du problème dans le cas de la reconnaissance de la parole :

Tout en poursuivant la reconnaissance de la phrase en cours, un système de compréhension de la parole doit pouvoir accepter trois types d'erreur.

a) Insertion d'un mot parasite :

A l'intérieur d'une phrase syntaxiquement correcte, peuvent apparaître des mots ne faisant pas partie de la définition du langage. Ce cas est assez fréquent dans la conversation orale ; il suffit pour s'en rendre compte de noter l'utilisation abusive de monosyllabes tels que "Euh", "bon", "bien"... L'adjonction d'éléments de ce type provoque une erreur au cours de la reconnaissance et pourtant, lors de leur rencontre, il suffit de progresser de quelques positions dans la chaîne phonémique pour pouvoir poursuivre normalement la reconnaissance. Il en est de

même lors de l'insertion d'un groupe de mots n'apportant aucune information sémantique supplémentaire.

Ainsi, si nous nous plaçons dans le cadre d'un langage oral d'interrogation d'une banque de donnée, les phrases suivantes devraient toutes être acceptées comme équivalentes même si seule la première est syntaxiquement correcte :

"quel est le nom de l'auteur de l'Illiade ?

"quel est le nom de l'auteur, Euh, de l'Illiade ?"

"quel est le nom, si vous pouvez me le dire, de l'auteur de l'Illiade ?"

b) Substitution :

Dans une langue possédant beaucoup de synonymes, il peut arriver qu'un mot soit remplacé par un autre, ne faisant pas partie du lexique du système de reconnaissance. C'est là encore une cause d'échec de la reconnaissance alors qu'il suffirait de considérer que le mot a été bien reconnu, si du moins il est possible de détecter dans la suite phonémique restant à traiter le début de son contexte droit.

Car pourquoi accepter une phrase telle que :

"Je voudrais la référence de..."

et rejeter :

"J'aimerais avoir la référence de..."

Il nous faut pourtant émettre quelques restrictions à ce genre de substitution : en aucun cas on ne peut les accepter sur des mots caractéristiques au niveau sémantique (ce que certains auteurs appellent des mots clés). En effet si on a la phrase "est ce que le chat est un herbivore" et que "chat" ne fait pas partie du lexique mais qu'on y trouve "cheval". Il ne faut pas pour autant considérer qu'il y ait eu substitution. Nous avons donc besoin ici d'une liste de mots caractéristiques au niveau sémantique telle que celle définie en partie B, paragraphe 2-2-3-.

c) élision :

Le dernier cas à considérer est celui de l'élision d'un mot ou groupe morphologique. Si nous reprenons l'exemple donné en (a), les phrases :

"quel est le nom de l'auteur de l'Illiade ?"

et "quel est l'auteur de l'Illiade ?"

sont sémantiquement équivalentes ; elles doivent donc être reconnues au même titre. Il suffit pour reconnaître la seconde de considérer que "le nom de" a été reconnu et d'avancer d'un pas dans l'analyse syntaxique.

A ce stade, la même restriction que celle proposée en (b) sera retenue : on n'acceptera pas d'élision sur des mots sémantiquement caractéristiques.

4-2-2- Solution adoptée : analyse syntaxique pas à pas :

L'analyseur lexicographique prend donc en compte les hypothèses H1 émises par le niveau syntaxique et tente la reconnaissance des terminaux correspondants. Si l'une au moins des hypothèses est validée par le système de reconnaissance de mot, le retour se fait aussitôt à l'analyseur syntaxique (cf § 4-1-2-). Dans le cas contraire l'étage lexicographique génère de nouvelles hypothèses, compte tenu des remarques faites au paragraphe précédent, et confirme ou infirme celles-ci avant de déterminer le point de reprise de l'analyse syntaxique suivant le principe exposé au paragraphe 4-1-.

a) génération de nouvelles hypothèses :

L'ensemble des nouvelles hypothèses, NH, doit contenir :

- (i) les anciennes hypothèses, H1, pour le cas où l'échec de la reconnaissance a été provoqué par l'insertion d'un mot parasite.
- (ii) les hypothèses formées des prochains terminaux à reconnaître, dans le cas d'élision ou de substitution d'un mot, ainsi que leur contexte d'analyse syntaxique. Notons H2 cet ensemble :

On a donc :

$$NH = H1 \cup H2$$

H1 est connu, il reste donc à déterminer H2.

H2 est composé des contextes droits des éléments de H1 non caractéristiques, au niveau sémantique. Soit H3 l'ensemble de ces éléments et C celui des mots caractéristiques fournis en données, on obtient :

$$H3 = H1 - C$$

Il ne reste plus qu'à déterminer l'ensemble des contextes droits de chaque élément e_i de H3 : notons h_i cet ensemble, h_i correspond aux hypothèses qui seraient émises par l'analyse syntaxique si celle-ci était poursuivie après la reconnaissance de e_i .

Ainsi, on détermine cet ensemble h_i en supposant e_i normalement reconnu et en poursuivant l'analyse syntaxique d'un pas.

Lorsque les h_i correspondant à tous les e_i sont construits, leur réunion forme l'ensemble H2 :

$$H2 = \bigcup_i h_i$$

c.à.d. $e_i \in H3$

On a donc l'algorithme suivant :

H1 = donnée ; C = donnée

H3 = H1 \cap C

H2 = \emptyset

Pour chaque $e_i \in H3$ considéré comme reconnu, faire

Analyseur syntaxique d'un pas : jusqu'à la détermination de H_i

$H2 = H2 \cup h_i$

$NH = H1 \cup H2$

b) vérification de ces hypothèses :

La vérification de ces hypothèses se fait elle aussi à l'aide du module de reconnaissance de mots (cf chapitre 5), mais le principe est différent de celui utilisé lors du traitement des hypothèses émises par le niveau syntaxique.

Lors du test des hypothèses H1, le mot recherché devait se trouver en position gauche de la partie de phrase non encore traitée. Ici, la chaîne phonémique correspondant au mot recherché peut être précédée d'une sous-suite phonémique, plus ou moins longue, correspondant à une insertion ou à une substitution possible. Il s'agit donc de sélectionner un mot dans cette chaîne plutôt que de le reconnaître en tête. Néanmoins nous utiliserons le même module de reconnaissance de mots et nous réaliserons la sélection au moyen d'une boucle faisant progresser l'indice de départ de la reconnaissance.

4-3- DESCRIPTION FONCTIONNELLE DE L'ANALYSEUR LEXICOGRAPHIQUE :

Les données utilisées dans l'analyseur lexicographique sont :

- l'ensemble des hypothèses émises par le niveau syntaxique; elles sont représentées dans un tableau TAB(*,3) où chaque élément est composé d'un triplet (numéro de terminal, et deux variables définissant le contexte de l'analyse syntaxique). On notera N, leur nombre.

- les informations définissant le langage de l'application : grammaire, lexicque, informations sémantiques, telles que nous les avons définies partie B paragraphe 2-2-.

- Enfin la zone de sauvegarde des retours arrière utilisée par l'analyseur syntaxique à laquelle il a accès.

La figure 4-1- donne un schéma général de cet analyseur lexicographique qui correspond à la partie 4 du schéma général du système donné en partie B paragraphe 1-4-2-.

L'algorithme correspondant à la réalisation de cet analyseur lexicographique peut se décrire ainsi :

- tri du tableau TAB , correspond aux hypothèses H1, d'après le numéro de terminal TAB (*,1)

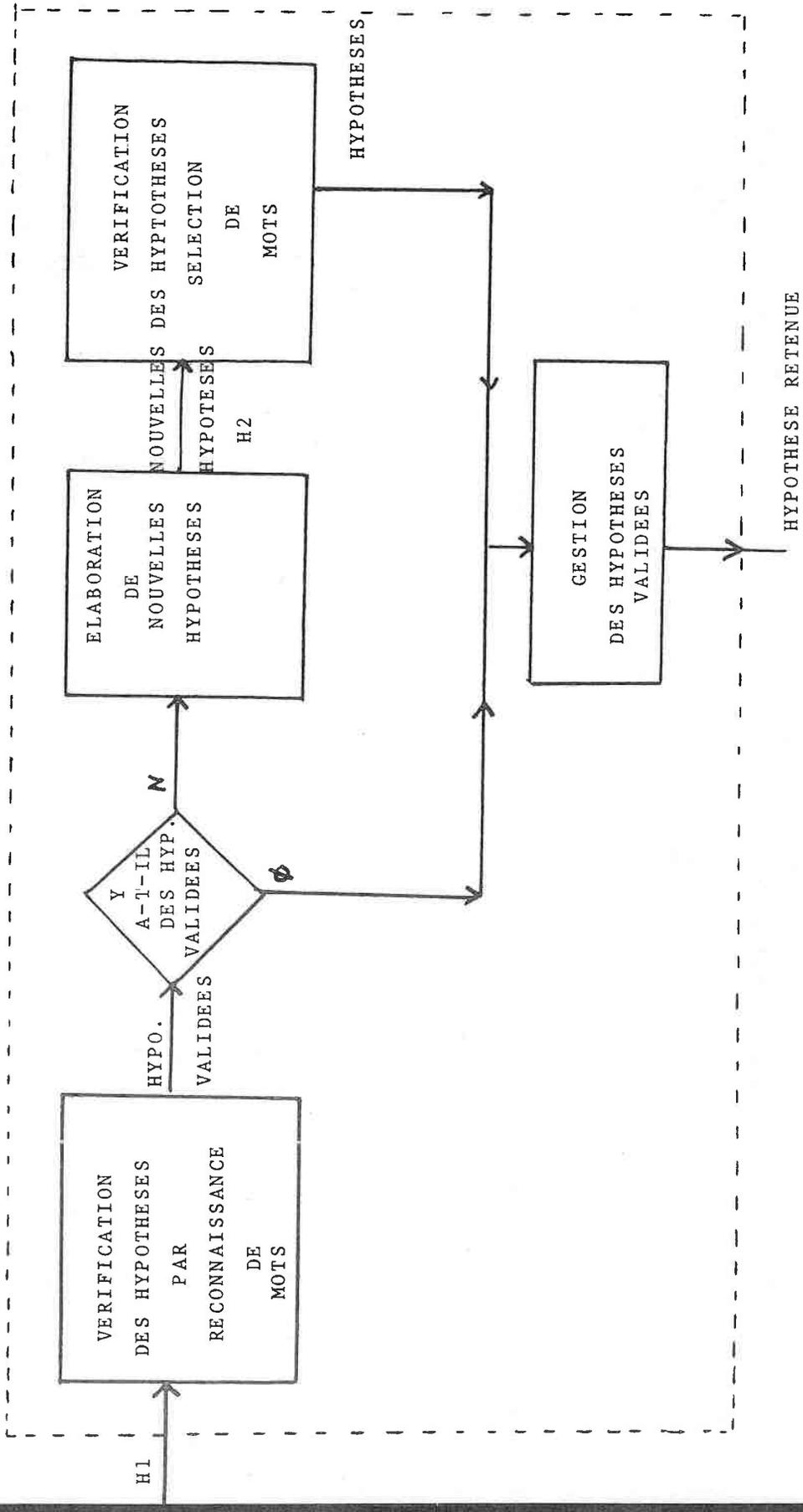


Schéma de l'analyseur lexicographique

figure 4-1-

- R = 0; N1=0 (N1 donne le nombre d'hypothèses validées)

- Pour I de 1 à N faire

si TAB(I,1)≠R alors début

-Appel du module de reconnaissance
de mot avec TAB(I,1) comme paramètre
résultat : IND=VRAI si reconnu
=FAUX sinon

- R=TAB(I,1)

fin

si IND alors début

N1=N1 + 1

rangement de l'Hypothèse correspondante dans
un tableau T1(*,5) des Hypothèses validées.

T1(N1,4)=taux de réussite S_i de la reconnaissance

T1(N1,5)=indice IP du dernier phonème traité dans
la phrase à reconnaître.

fin

- si N1 = 0 alors début

****** cas où aucune hypothèse n'a été validée.

- détermination de l'ensemble H3 correspondant
aux hypothèses non caractéristiques sémantique-
ment et rangement dans un tableau, trié,
HYP3 ; NB3 = nombre d'élément de HYP3

- Pour I de 1 à NB3, faire

- restauration du contexte d'analyse syntaxique
de l'élément I de HYP3

- poursuite de l'analyse syntaxique à partir
du point de reprise ainsi déterminé jusqu'à
l'élaboration de NI nouvelles hypothèses HI.

- si NI≠0 alors début

- insertion de ces nouvelles hypothèses HI
dans le tableau TAB trié des Hypothèses H1
et progression de N

fin

- R = 0

- Pour I de 1 à N, faire

si TAB(I,1) ≠ R alors début

sélection du mot correspondant au terminal
TAB(I,1).

* Cela correspond à une boucle sur l'appel
* du module de reconnaissance de mot
* en faisant progresser l'indice phonémique
* de départ dans la phrase à reconnaître.
* En cas d'échec un indicateur IND est mis
* à FAUX sinon il est positionné à VRAI

fin

si IND=VRAI alors début

- rangement de la (ou les) hypothèse (s)
ainsi validé(es) dans le tableau T1
- et mise à jour de N1

fin

fin

- Tri du tableau T1 en fonction des taux de réussite S_i
croissant, et transfert sur la pile des Hypothèses
validées.

- si Pile non vide alors début

détermination du point de reprise de l'Analyse
syntaxique en fonction des informations fournies
par le sommet de pile : (NTER, J, IR, S_i , IP)

fin

sinon NTER = 0

fin du module d'analyse lexicographique.

C H A P I T R E 5

LE MODULE DE RECONNAISSANCE DE MOTS

5-1- INTRODUCTION :

Avec le module de reconnaissance de mots, nous abordons la partie de reconnaissance proprement dite. Comme nous l'avons déjà noté (partie A paragraphe 2-3), il existe deux approches nettement distinctes de la reconnaissance vocale :

- la reconnaissance "analytique" ou par "sons élémentaires"
- la reconnaissance "globale" ou par mot.

Pour notre part, c'est dans le cadre de la reconnaissance analytique que nous travaillons et le module de reconnaissance de mots, dont nous avons besoin, consiste en un algorithme de recherche lexicale à partir de la chaîne phonémique de la phrase à reconnaître.

Nous verrons successivement dans ce chapitre :

- a) l'étude des principales difficultés à traiter lors de la recherche lexicale dans le cadre de la reconnaissance du discours continu.
- b) les diverses méthodes de recherches lexicales.
- c) enfin la méthode utilisée dans notre système, sa réalisation et les résultats qu'elle fournit.

5-2- PRINCIPALES DIFFICULTES :

La reconnaissance phonémique de mots ou de phrases soulève de multiples difficultés qui proviennent soit de problèmes purement phonétiques (2) que l'on traite habituellement en "phonétique combinatoire", soit de caractéristiques propres à la reconnaissance automatique de la parole (36). Après le traitement acoustique, la transcription phonémique de la phrase à reconnaître reste fortement entachée d'erreurs. Il est, certes, intéressant d'avoir une chaîne phonémique aussi bonne que

possible, car sa qualité détermine, pour une grande mesure, les performances des recherches lexicales. Mais ici, la solution idéale n'existe pas, car la recherche lexicale est rendue délicate par un certain nombre de facteurs que l'on peut classer en quatre groupes, du fait que :

- les frontières de mots ne sont pas apparentes
- des fusions phonétiques sont souvent possibles
- des altérations phonologiques peuvent apparaître
- des erreurs peuvent être provoquées par les limites du système acoustique.

a) Les frontières de mots ne sont pas apparentes :

Comme nous l'avons déjà noté (partie A, paragraphe 1-2-2-) l'une des caractéristiques de la parole est la continuité, et il en est de même pour sa transcription phonémique. Ce problème se trouve accentué par certaines modifications apparaissant aux frontières de mots. En effet, la prononciation correcte n'est pas unique et plusieurs chaînes différentes peuvent représenter la même phrase suivant qu'il y a ou non des liaisons ou des élisions.

- les liaisons se font en général entre le phonème terminal d'un mot et le premier du mot suivant. Parmi elles, certaines sont obligatoires, d'autres non.

Exemples : on dit : "un enfant" et non pas "un/enfant" ;
par contre on peut dire aussi bien
"ils courent/encore que "ils courent encore".

- les élisions en fin de mot sont aussi très courantes spécialement lorsque le dernier phonème correspondant à un e muet (ə).

Certaines existent jusque dans l'écriture : c'est le cas de "le" qui devient "l'" devant une voyelle.

D'autres apparaissent plus rarement (par exemple lorsque le e muet est compris entre deux consonnes), et sont en général provoquées par une diction trop rapide.

Exemple : j'voudrais" pour "je voudrais".

D'autres enfin se combinent avec une liaison possible après que le "e" muet soit tombé.

Exemple : "soupe épaisse" peut devenir "soup'/'épaisse" ou "soup' épaisse".

- A cela s'ajoutent les problèmes particuliers posés par les pluriels. La règle générale est d'ajouter un s en fin de mot, mais il existe aussi diverses règles particulières telles que :

AL → AUX

EU → EUX

Phonétiquement, cette marque du pluriel n'a, le plus souvent, aucune incidence ; elle fait pourtant apparaître une nouvelle liaison possible :

Ex : si nous reprenons l'exemple précédent on peut obtenir :

"soupes épaisses"

"soup 's épaisses"

"soup'/'épaisses"

voir même "soup' épaisses"....

b) des fusions phonétiques (ou assimilations) sont souvent possibles :

Les chercheurs américains ont mis l'accent sur de nombreuses fusions phonétiques possibles dans leur langue posant de gros problèmes lors de la reconnaissance : la phrase "won't you" peut se transcrire (wount ju) ou (wountf) ; (t j) devient ainsi (t ʃ) ; il ne faut pas croire qu'il en soit

autrement en français ; pour s'en convaincre il suffit de prononcer un peu rapidement "une heure et demie" pour s'apercevoir que l'on dit en fait "une heure é-n-mi" : au contact de la nasale m, le d s'est nasalisé pour devenir n.

c) des altérations phonologiques peuvent apparaître :

Ces altérations peuvent provenir soit d'une mauvaise prononciation soit d'un effet du contexte sur la partie modifiée.

On peut y distinguer :

- des réductions de phonèmes :

ex : (Y) → (ə) dans "te vois" au lieu de "tu vois"

- des omissions de voyelles : "donn(e)rai" pour "donnerai"

de consonnes : "quē(1)que" chose" pour "quelque chose"

voir même des deux : on dira "tab(le)d'honneur" pour "table d'Honneur"

- des insertions, le plus souvent de voyelles, pour faciliter la prononciation :

"arque de triomphe" "ourse blanc".

d) d'autres altérations peuvent être prononcées par les limites du système acoustique. (partie B paragraphe 1-3-2).

Ce sont par exemple (1) :

- la mauvaise distinction entre les plosives non voisées et voisées

- les confusions entre fricatives

- l'omission de consonnes plosives en début de mot

- les confusions entre phonèmes introduites par le type d'analyse acoustique utilisée.

Dans le cadre d'une recherche lexicale, il n'est pas possible de traiter chaque cas de manière particulière. Néanmoins, afin de rendre compte, le mieux possible, de ses diverses variations, nous les regrouperons en trois grands types (36):

- 1- substitution de phonèmes.
- 2- omission de phonèmes.
- 3- inversion de phonèmes.

De plus nous tiendrons compte de la possibilité d'avoir des répétitions d'un même phonème, ou de phonèmes voisins (ceci correspond à un phénomène de segmentation multiple).

5-3- LES METHODES DE RECHERCHE LEXICALE :

Les travaux dans ce domaine sont encore en nombre relativement restreint, dans le contexte général de la reconnaissance vocale, car les problèmes de reconnaissance analytique du discours continu ne sont véritablement abordés que depuis peu de temps (partie A ch. 2).

Une méthode de "comparaison élastique" a été proposée, tenant compte à la fois des éléments composant la chaîne et de leur ordre (56). Il semble qu'elle puisse être utilisée dans le cas de la reconnaissance analytique de la parole bien qu'elle ait été présentée dans le cadre plus simple de correction d'erreurs dans un texte dactylographié.

Une seconde méthode est fondée sur une procédure d'extraction de sous-chaînes ordonnées de longueur maximale commune à deux chaînes d'éléments. De là on obtient une méthode permettant le calcul d'un indice de ressemblance entre deux chaînes de phonèmes et acceptant des réponses multiples pour chaque élément de la chaîne (57).

Une autre technique de décodage, fondée sur l'algorithme de FANO (58) ouvre des perspectives intéressantes dans le domaine de la parole, moyennant certaines transformations. Il s'agit essentiellement d'une procédure de recherche de chemin dans un graphe avec une métrique liée à la ressemblance entre une chaîne d'éléments donnée et la chaîne correspondant à ce chemin. Une application a été proposée dans le cas de la parole (59) ; elle est alors utilisée avec une analyse syntaxique pour la reconnaissance de phrase (36) et rejoint par là le travail d'ALTER sur le Fortran (24).

Une quatrième méthode, dont le but est identique à celle de FANO, a été présentée par J.P. HATON (1) : Il s'agit de chercher un chemin dans une sorte de graphe obtenu en considérant des réponses multiples dans la chaîne phonémique. Le nombre de phonèmes a été limité à trois. Chaque mot du lexique est décrit sous forme d'un arbre donnant sa transcription phonétique avec éventuellement ses variantes possibles. Chaque transition d'un phonème à un autre possède un poids déterminé par la forme de la représentation lexicale du mot considéré et par un certain nombre de règles de transitions. La comparaison d'une chaîne phonémique avec un mot du lexique revient ainsi à rechercher le chemin optimal (celui qui a le plus grand poids) permettant de décrire cette chaîne. Plusieurs chemins sont possibles, et parmi eux certains sont abandonnés en cours de route grâce à des tests d'impossibilité.

Une autre méthode proposée par J.P. HATON utilise le principe de comparaison dynamique, déjà utilisé lors de la reconnaissance globale de mots (44), (45). Il s'agit de minimiser le taux de dissemblance entre deux chaînes, au sens d'une métrique interphonème. Cette métrique

utilise les données d'une matrice de proximité de phonèmes obtenue par apprentissage sur le système acoustique, de façon analogue à une matrice de confusion. Cette façon de procéder permet de tenir compte le mieux possible des contraintes phonétiques: la distance d ainsi déterminée dépend de la proximité phonétique des phonèmes et des performances du reconnaiseur acoustique.

La figure B-5-1- reprend un exemple donné dans (1). On a représenté le résultat de la comparaison de la chaîne phonémique correspondant à la prononciation du mot "SAPIN", avec la représentation lexicale de ce même mot ($/sa p \hat{E}/$) qui été finalement reconnu.

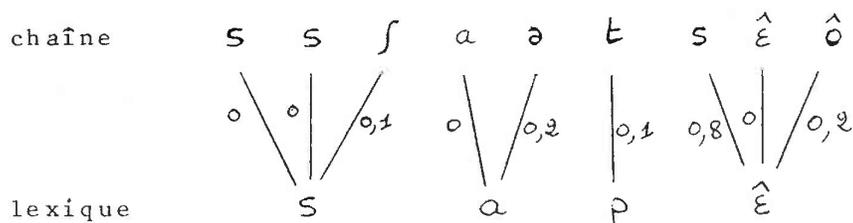


figure B-5-1-

Comparaison d'une chaîne inconnue avec le mot sapin.

Le temps de comparaison d'une chaîne inconnue avec un mot de référence varie en moyenne de 1 à 2 ms, selon la longueur, sur un mini calculateur T 2000, relativement peu rapide. Ces performances permettent l'utilisation de vocabulaires importants (plusieurs centaines de mots), avec réponse quasi-immédiates, d'autant plus qu'il est possible d'envisager des méthodes de position du lexique en fonction de certains critères phonétiques.

Toutes les méthodes que nous avons vues ici sont des méthodes de recherche lexicale pure et répondent toutes à la question : étant donné une chaîne phonémique inconnue quel mot du lexique représente-t-elle ? Pour notre part nous posons le problème de manière quelque peu différente : étant donné les hypothèses émises par le niveau syntaxique quelles sont celles qui peuvent être validées ? Il s'agit moins d'une reconnaissance que d'une vérification du bien fondé de telle ou telle hypothèse. Mais en fait, la reconnaissance et la vérification sont deux problèmes très proches.

5-4- PRESENTATION DE LA SOLUTION MISE EN OEUVRE :

La solution que nous avons adoptée est proche à la fois de la méthode utilisée par C. TAPPERT (36), qui est une application de l'algorithme de FANO, et de celle de cheminement dans un tableau proposée par J.P. HATON (1). Elle consiste à vérifier s'il existe un chemin possible dans le graphe formé par la chaîne phonémique à reconnaître et décrivant une représentation de la chaîne de référence du terminal à reconnaître. En cas de réussite, elle considère l'hypothèse ainsi testée comme validée et lui affecte un score S_i utilisé à un niveau supérieur par l'analyseur syntaxique (porte B chapitre 3).

Comme nous l'avons déjà vu partie B chapitre 2, la phrase à reconnaître peut être représentée par un pseudo-graphe, GRAPH (LG,3) tenant compte pour chaque phonème des possibilités les plus probables ; chaque mot de référence, terminal vrai ou terminal de type lexicale, est par ailleurs décrit dans un dictionnaire sous forme d'un arbre donnant sa transcription phonétique, avec éventuellement les variantes possibles : élision, substitution, insertion de phonèmes (cf. figures B-2-2- et B-2-7-).

Il s'agit donc d'essayer de décrire complètement l'une des branches du graphe représentant le mot à tester au moyen d'un chemin dans le tableau GRAPH. Chaque chemin ainsi parcouru est affecté d'un score de réussite S défini par la relation récurrente suivante :

Soit I le nombre de phonèmes traités dans la chaîne à reconnaître.

TD_I le taux de dissemblance cumulé après traitement de I phonèmes correspondant au mot testé.

TDP_I le taux de dissemblance entre le phonème I de la chaîne et celui qui est supposé lui correspondre dans le graphe de référence

$$\text{et } 0 \leq TD_I \leq 1 ; \quad 0 \leq S_I \leq 1 ; \quad 0 \leq TDP_I \leq 1$$

$$\text{alors } S_0 = 1 ; \quad TD_0 = 0$$

$$TD_I = TD_{I-1} + TDP_I$$

$$S_I = 1 - \frac{2 \times TD_I}{I}$$

La comparaison revient à rechercher le chemin optimal -de taux de réussite le plus fort- pour le terminal à traiter. En fait, nous avons été amenés ici à distinguer le traitement des terminaux vrais et celui des terminaux de type lexique. Ces derniers peuvent en effet avoir plusieurs réalisations différentes, il faut donc poursuivre tous les chemins menant à l'une de ces réalisations et accepter un résultat composé de plusieurs couples (réalisation K, taux de réussite associé).

De plus, chaque fois qu'un chemin acquiert un score de réussite inférieur à un certain seuil, déterminé de manière expérimentale par essais successifs, ($S_I < \text{SEUIL}$), la vérification s'arrête ; ce chemin est alors considéré comme erroné et, dans le cas de terminaux de type lexique, le processus se poursuit sur un autre chemin non encore exploré. Il serait également intéressant de prévoir de tels abandons sur d'autres critères tels que : absence ou présence de fricatives ou d'occlusions ; mais jusqu'alors de tels tests ne sont pas encore implémentés dans notre système.

Lors de la comparaison, le chemin suivi dans le tableau GRAPH et dans le graphe de référence du terminal à tester sera le plus possible de type continu, compte tenu des substitutions, élisions ou insertions prévues dans le lexique. Le taux de dissemblance TDP_I sera mis à zéro, sauf dans les cas d'élision ou d'insertion possibles, où il prendra la valeur TEI. De plus, on acceptera dans la chaîne d'entrée des répétitions d'un même phonème consécutives à des segmentations multiples. Chacune d'entre elles augmentera le taux de dissemblance de TR.

Si, malgré tout, au cours du traitement, un chemin continu ne mène à rien et rend impossible la reconnaissance du phonème I, le système émet trois hypothèses correspondant toutes à des cas non prévus dans la chaîne de référence (cf paragraphe 5-2).

- (i) hypothèse de substitution : le système considère qu'un phonème en a remplacé un autre et provoque alors la progression d'un pas (correspondant à un phonème) tant dans la chaîne d'entrée que dans celle de référence.
- (ii) hypothèse d'élision : il considère qu'un phonème manque dans la chaîne à reconnaître et fait progresser d'un pas l'indice dans le graphe de référence du dictionnaire.

(iii) hypothèse d'insertion : il considère qu'un phonème parasite s'est introduit dans la chaîne d'entrée et fait progresser d'un pas l'indice de traitement dans cette chaîne.

Le système vérifie ces diverses hypothèses et, si l'une d'elles s'avère justifiée (possibilité de traiter alors le phonème suivant) il poursuit la reconnaissance en augmentant préalablement le taux de dissemblance de TDP_I d'une valeur égale à $TSUB$, $TELI$ ou $TINS$ suivant l'hypothèse prise en compte.

Les différentes valeurs de TDP_I : TEI , TR , $TSUB$, $TELI$ et $TINS$ seront déterminées pour chaque application de manière expérimentale par tests successifs. Elles déterminent, pour une grande partie, le taux de mauvaise reconnaissance ou de non reconnaissance du système : plus ces valeurs seront importantes, plus la reconnaissance sera stricte et provoquera le rejet des phrases mal prononcées.

Il reste enfin à traiter le problème de jonction de mots que nous avons présenté au paragraphe 5-2. Nous avons décidé, pour en rendre compte, de ne pas traiter le dernier phonème correspondant à la représentation lexicale d'un mot dans le cas où celui-ci peut être l'objet d'une élision. Nous utilisons donc une variable $ELIS$ prenant pour valeur celle du dernier phonème du mot traité suivant qu'il y a élision ou pas. La présence de ce phonème sera testée au début de la reconnaissance du mot suivant grâce à $ELIS$. Ce n'est là qu'une simple ébauche du traitement des jonctions de mots ; il faudrait en fait tenir compte de nombreuses règles faisant intervenir les pluriels, les liaisons, les élisions et, à un autre niveau, les conjugaisons ou les déclinaisons (60).

Le système de reconnaissance de mots ainsi réalisé nous a donné des résultats intéressants dans le cadre de notre système et nous permet de travailler en temps réel dans le cas où le nombre d'Hypothèses à vérifier à chaque pas de l'analyse syntaxique n'est pas trop important : de l'ordre de quelques dizaines. En fait, il ne résoud pas tous les problèmes de recherche lexicale rencontrés en reconnaissance du discours continu. Une fois de plus nous pouvons dire qu'il représente une étape et non un aboutissement : il doit être possible de l'optimiser en temps de calcul comme en efficacité de reconnaissance, et il serait sans doute intéressant de pouvoir faire un test comparé des différentes méthodes de recherche lexicale utilisées jusqu'alors, bien qu'elles aient été conçues pour la plupart dans des contextes différents.

5-5- DESCRIPTION FONCTIONNELLE DU MODULE DE RECONNAISSANCE DE MOTS:

Le module de reconnaissance de mots comporte deux parties, l'une correspondant aux terminaux vrais l'autre aux terminaux de type lexicale. Le principe de ces 2 parties est le même, seule la réalisation diffère en raison de représentations différentes des chaînes de référence permettant d'accroître l'efficacité dans l'un et l'autre cas. Dans ce paragraphe, nous ne décrivons que la première partie afin de mieux voir comment sont gérées les diverses hypothèses, dont nous avons parlé au paragraphe précédent, sans pour autant alourdir la présentation par des considérations trop proches de la programmation.

Auparavant, voici quelques rappels sur les notations utilisées.

- Pour les chaînes phonémiques, on utilise les représentations données dans le chapitre 2 (partie B) :

Un mot M lexicale est déterminé par une partie du tableau $TER(LT,3)$ commençant à l'indice $TERMI(M)$.

la chaîne phonémique, quant à elle, est définie par un tableau $GRAPH(LG,3)$.

- I représente l'indice courant dans le tableau GRAPH
- K celui correspondant dans le tableau TER
- AC donne la valeur du dernier phonème traité pouvant être répété dans la chaîne.
- l'élément GRAPH(I,*) représente indistinctement GRAPH (I, 1), GRAPH(I,2) ou GRAPH(I,3).
- ELIS indique le numéro d'un phonème traité qui correspond à une élision possible à la fin du mot précédent.

Etant donné ces notations, l'algorithme utilisé peut se décrire ainsi :

* Initialisation :

K = TERMI(M)

I = PT ;

AC = ELIS

TD = TDI = 0 ;

* Boucle de reconnaissance

tant que TER(K,1) > 0 faire

traitement du phonème suivant

* résultat : quadruplet (M,I,S,ELIS)

ELIS = - TER(K,1)

$$S = 1 - \frac{2 \times TD}{I-PT}$$

Le traitement du phonème suivant correspond à la prise en compte des diverses possibilités exposées au paragraphe 5-4- Huit cas peuvent se présenter ; ils sont regroupés dans une table de décision (fig. B-5-5-), et seront traités de manière séquentielle et ordonnée.

Condition	Action correspondante
<u>Cas Idéal</u> $\text{GRAPH}(I,*) = \text{TER}(K,1)$ ou $\text{GRAPH}(I,*) = \text{TER}(K,2)$	$I=I+1$ $K=K+1$ $AC=\text{GRAPH}(I,*)$
<u>Insertion prévue</u> $\text{CRAPH}(I,*) = \text{TER}(K,3)$	$I=I+1$ $AC=\text{TER}(K,3)$ $\text{TDP}=\text{TDP}+\text{TEI}$
<u>Elision prévue</u> $\text{TER}(K,1) \quad 0$	$K=K+1$ $AC=0$ $\text{TDP}=\text{TDP}+\text{TEI}$
<u>Répétition du phonème prec.</u> $\text{CRAPH}(I,*) = AC$	$I=I+1$ $\text{TDP}=\text{TDP}+\text{TR}$
<u>Hypothèse d'élision</u> $G = 0$	$AC=0$ $\text{TDP}=\text{TDP}+\text{TELI}$ $K=K+1$
<u>Hypothèse de substitution</u> $G = 1$	$I=I+1$ $\text{TDP}=\text{TDP}+\text{TSUB}$
<u>Hypothèse d'Insertion</u> $G = 2$	$K=K-1$ $\text{TDP}=\text{TDP}+\text{TINS}$
Autres cas	erreur de reconnaissance $R = 0$ fin de la reconnaissance

$G = 0$
 $\text{TD}=\text{TD}+\text{TDP}$
si $(\text{SEUIL} > \frac{1-2 \times \text{TD}}{I-\text{PT}})$

alors

erreur de reconnaissance

$R = 0$

fin de la reconnaissance

$G = G + 1$

figure B-5-5-

5-6- EXEMPLES DE RECONNAISSANCE :

A titre d'exemple, nous avons regroupé dans ce paragraphe, des chemins suivis lors de la reconnaissance de différents mots et le taux de réussite qui leur était affecté.

Au cours de ces tests, nous avons fixé les différents taux de dissemblance ainsi :

$$TEI = TR = 0.3$$

$$TINS = TSUB = TELIS = 0.5$$

le seuil d'abandon, quant à lui, avait pour valeur 0.6.

Dans la suite, les pointeurs A et B représenteront respectivement l'indice de départ et l'indice de fin de reconnaissance dans la chaîne phonémique à traiter.

5-6-1- CAS DE TERMINAUX VRAIS :

Prenons, par exemple, le cas de la recherche du terminal "Je voudrais" dans diverses chaînes phonémiques ; l'algorithme recherche donc le chemin optimum décrivant la chaîne de référence correspondant à ce mot.

La figure B-5-6-1- donne sous forme de graphe la représentation lexicale "Je voudrais".

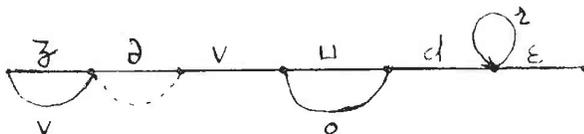


figure B-5-6-1-

Voici les résultats obtenus pour trois chaînes phonémiques différentes :

a) ... a l u [z] a z γ [u] b i ε p e l e ...
 ... e z o v [ə] [v] [u] e [d] e e t a r ε ...
 ... ə o z e z o g [ε] i k ə i ...
 ... Allô Je voudrais parler ...
 A B

On peut remarquer ici que l'algorithme a sélectionné une chaîne correspondant exactement à la représentation la plus probable au vu du lexique, mise à part une répétition du /u/ de "voudrais".

Le taux de réussite : $S=0,91$ correspond à $S=1-\frac{2 \times TR}{2}$

b) a l o z [u] [d] [ε] l p o s
 e z o v γ b i z t o f ...
 ə u [z] o g e e k a ...
Allô j'voudrais l' post'...
 A B

Dans ce deuxième exemple, correspondant à une diction très rapide, l'élision du /ə/ de "je" a été prise en compte. De plus le système a émis une hypothèse d'élision sur le de "voudrais" qui s'est vérifiée par la bonne reconnaissance de la fin du mot.

Le taux de réussite : $S = 0.6$ correspond bien à

$$S=1-\frac{2*(TEI + TELI)}{4}$$

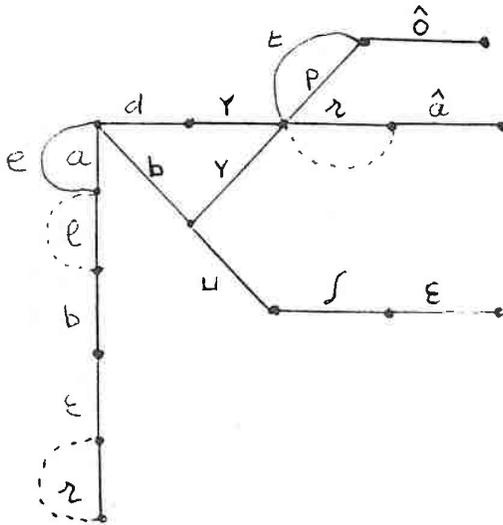


figure B-5-6-2-

La phrase à reconnaître correspond à la chaîne phonémique suivante :

m	ə	s	ə	b	γ	p	ô	m	e	l	s	ε
b	e	s	e	d	u	t	â	b	a	r	f	i
d	a	f	g	k	d	ə	s					
	↑	↑	↑	↑	↑	↑	↑					
monsieur			A	Dupond		Merci						

Rappelons que dans ce cas l'algorithme recherche tous les chemins décrivant complètement une réalisation du terminal à traiter.

Le résultat de la reconnaissance est double : Dupond et Durand sont tous deux reconnus.

Dupond avec un taux $S=1$ par deux chemins équivalents :

...	ə	b	γ	p	ô	m	ə	b	γ	p	ô	m	...	
...	e	d	u	t	â	b	...	ET	...	e	d	u	t	â	b	...
...		g	k	d	g	k	d		
...	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
...	A		B		A		B		A		B		A		B	

C H A P I T R E 6

RESULTATS FOURNIS PAR LE SYSTEME

6-1- PHRASE BIEN RECONNUE ET DIFFERENTS TYPES D'AMBIGUITES :

Les résultats fournis par le système après une tentative de reconnaissance d'une phrase sont de deux types : échec ou réussite.

a) Le cas de l'échec correspond dans notre système à un renvoi à une procédure d'erreur ERR (N) où trois cas sont à distinguer :

(i) N = 1 : à un moment donné, au cours de la reconnaissance et après traitement de l'analyseur lexicographique, aucune des hypothèses émises par le niveau syntaxique n'est validée,

(ii) N = 2 : l'analyseur syntaxique détecte la fin de la reconnaissance avant que toute la chaîne d'entrée ne soit traitée,

(iii) N = 3 : la chaîne d'entrée est reconnue entièrement, mais elle est syntaxiquement inachevée.

Nous avons opté pour la non-formalisation du traitement de ces erreurs car, si le plus souvent on peut supposer que la solution consisterait à provoquer une demande de répétition, rien n'empêche, dans le cas d'applications spécifiques, de prévoir d'autres possibilités. Nous considérons donc que la procédure ERR (N) est à redéfinir et à ré-écrire pour chaque application, compte tenu de ces trois types d'erreur, et, éventuellement, pour d'autres types d'erreur propres à l'application.

b) En cas de réussite, le résultat fourni est composé de trois éléments :

- la phrase reconnue, codifiée dans un tableau de IR éléments.
- le taux de réussite cumulé SC associé à cette phrase.
- la pile des hypothèses validées par l'analyseur lexicographique et non traitées par le niveau syntaxique.

En fait, il nous faut distinguer à ce stade les phrases dont la reconnaissance est certaine et celles qui ont une ou plusieurs causes d'ambiguïté. On peut en effet distinguer trois causes d'ambiguïté pour une phrase reconnue :

(i) un taux de réussite SC trop faible. Suivant la valeur des paramètres utilisés pour le calculer (TEI, TR, TINS, TSUB, TELI, SEUIL...), il indique si la reconnaissance a été stricte ou au contraire très large (cf paragraphes B-3-3-3- et B-5-4).

(ii) un recours aux hypothèses supplémentaires émises par l'analyseur lexicographique (paragraphe B-4-2-2-). En effet leur utilisation s'accompagne d'une liberté croissante par rapport à la grammaire. On ne peut alors qu'émettre des suppositions sur le sens du message reconnu, sans pour autant avoir de certitude. Cette considération nous a conduit à utiliser un indicateur qui est positionné par l'utilisation de cette possibilité.

(iii) l'abandon de certains chemins au cours de l'analyse syntaxique. Le système de reconnaissance tel que nous l'avons développé travaille en se fondant à chaque instant sur le contexte gauche de la phrase ayant le plus fort taux de reconnaissance. Il parcourt ainsi le chemin le plus probable en laissant momentanément de côté d'autres possibilités. Par suite, si à la fin de la reconnaissance, certains chemins, n'ayant pas conduit à un échec, n'ont pas été entièrement parcourus, on ne peut pas dire que la reconnaissance effectuée est certaine mais simplement qu'elle est la plus probable.

Une fois de plus, on est confronté à un problème dont la solution ne semble pas pouvoir être formalisée : "comment lever ces ambiguïtés ?". Pour notre part, nous pensons qu'elles ne peuvent être levées efficacement que dans le cadre de chaque application. La figure B-6-1- schématise le système ainsi obtenu.

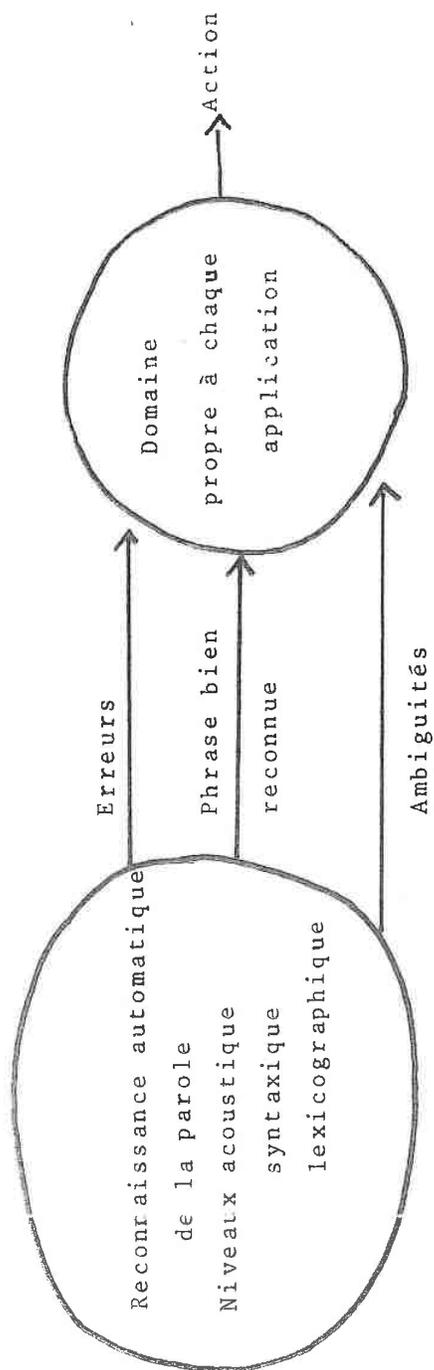


Figure B-6-1-

6-2- NECESSITE D'UNE PROCEDURE DE DIALOGUE PROPRE A CHAQUE APPLICATION

Le traitement des erreurs et des ambiguïtés propre à chaque application, sera essentiellement composé d'une procédure de dialogue permettant par le jeu des questions et des réponses de récupérer certaines erreurs ou de lever les ambiguïtés.

De telles procédures de dialogue ont déjà été proposées dans d'autres systèmes. C'est le cas principalement dans les travaux américains en cours de réalisation dans le cadre du projet ARPA. En France, une étude faite au CNET -Lannion (51) analyse différentes façons de tirer parti du dialogue pour l'amélioration de la reconnaissance automatique de la parole et pour celle de la communication homme-machine, et donne quelques enseignements apportés par une expérience de simulation d'un tel dialogue.

Dans le cadre de notre système, nous pensons que toute la phase de traitement propre à chaque application peut être orientée vers un dialogue entre la machine et l'utilisateur du système de reconnaissance.

Actuellement, comme nous le verrons en partie C, nous avons opté pour une simulation de cet échange, car nous ne possédons pas de sortie parlée. Il s'agit là d'un problème de moyens matériels et non d'un problème théorique qui n'enlève rien à la validité de nos résultats ; il existe en effet déjà des unités à réponse vocale pour ordinateur, appelées le plus souvent synthétiseurs. En France, plusieurs équipes travaillent sur ce sujet, et ont obtenu des résultats intéressants (61), (62), (63), (3), 64).

Le système présenté dans ce mémoire est ainsi appelé à travailler en lien avec un organe de sortie (synthétiseur, écran cathodique) permettant un dialogue fructueux entre l'homme et la machine. Son rôle est de valider les résultats de la reconnaissance avant de provoquer l'action correspondant à la demande émise dans le message oral ; cette procédure permet de réduire les risques inhérents à une fausse interprétation du message (par exemple en cas de commande d'une machine), et de limiter la nécessité de recours à un opérateur humain.

P A R T I E C

APPLICATION A UN LANGAGE DONNE

INTRODUCTION A LA PARTIE C :

Le système présenté dans la partie précédente de ce mémoire est un système général de reconnaissance du discours continu capable de traiter tous les langages définis par une C-grammaire. Cette dernière partie présente l'application mise en oeuvre dans le cadre d'un langage particulier : langage de communication avec un standard téléphonique (65). Cette application, choisie parmi les nombreuses possibles, a pour but de valider l'approche générale et de servir de support aux tests du système complet.

Nous verrons successivement les critères ayant conduit au choix de ce langage et sa définition, avant d'aborder la réalisation pratique, et plus particulièrement la procédure de dialogue propre à cette application. Les résultats obtenus en temps différé sur l'ordinateur IRIS 80 de l'IUCA de Nancy nous permettront de présenter les performances actuelles d'un tel système.

Enfin, un dernier chapitre est consacré à la poursuite possible de notre travail dans divers domaines : implémentation en temps réel, généralisation à plusieurs locuteurs et à un lexique plus important, apprentissage tant syntaxique que morphologique ou sémantique et définition de langages oraux.

C H A P I T R E 1

DEFINITION DU LANGAGE

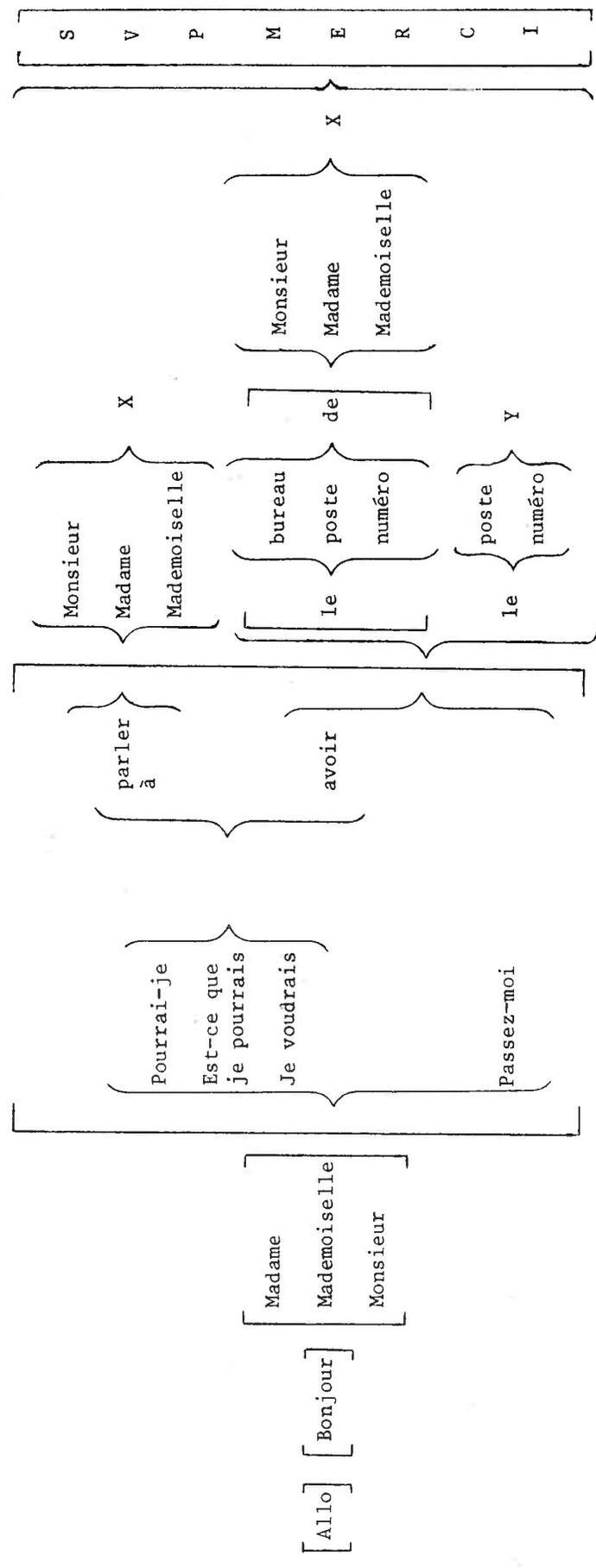
1-1- CHOIX DU LANGAGE : LANGAGE DE COMMUNICATION AVEC UN
STANDARD TELEPHONIQUE :

Comme nous l'avons vu dans la partie A chapitre 3, le domaine d'application de la reconnaissance de la parole est très vaste. Parmi les multiples possibilités de langages qui s'offraient à nous, il nous a fallu en sélectionner une comme support des différents tests du système permettant de valider notre approche.

Nous avons opté pour une application "grand public" utilisant le téléphone comme moyen de communication avec le système automatique. Certes, il ne s'agit pour le moment que d'une simulation ; mais une application de ce genre, outre son attrait, nous ouvre des possibilités intéressantes de poursuite de ce travail lors de généralisations ultérieures. Un tel système doit en effet pouvoir accepter un très grand nombre de locuteurs, hommes, femmes et enfants, bien qu'actuellement les résultats obtenus ne concernent qu'un locuteur unique ou une classe réduite de locuteurs dont les voix sont suffisamment proches.

Parmi ces applications "de type téléphonique", nous avons déjà cité la transmission de parole à faible débit (historiquement une des premières motivations de chercheurs dans le domaine, comme nous l'avons vu partie A chapitre 2), la consultation orale d'une base de données (centre de renseignements, fichiers, réservation de place, etc...) : chaque poste téléphonique est alors un terminal d'un ordinateur central, ou encore l'automatisation d'un standard téléphonique. C'est à cette dernière application que l'on s'intéresse ici. Après "espionnage" d'un standard téléphonique d'une entreprise, on a déduit un langage, sous-ensemble du français parlé, permettant de représenter une grande partie des phrases prononcées par un utilisateur s'adressant à une standardiste pour lui demander un correspondant.

La figure C-1-1- donne un aperçu général du langage de base utilisé.



{ } indique un choix possible entre plusieurs alternatives

[] indique la possibilité d'absence du groupe de mots correspondant

Aperçu général du langage utilisé

Figure C.1.1.

1-2- DEFINITION SYNTAXIQUE :

1-2-1- Grammaire :

Le langage utilisé comporte des phrases à structure assez souple dont la syntaxe de base est donnée par une grammaire à contexte libre. Cette grammaire est représentée dans le formalisme de Backus Naur sur la figure C-1-2-.

Le nombre de phrases correspondant à une syntaxe rigoureuse s'élève à plus de 7000 fois le nombre d'abonnés au standard. En fait, le nombre total de phrases, pouvant être reconnues par notre système avec cette grammaire, est infini car, comme nous l'avons vu en partie B chapitre 4, les analyseurs syntaxique et lexicographique utilisés tiennent compte d'élisions, de substitutions ou d'insertions possibles de mots quelconques dans ces phrases.

Ainsi à partir d'une phrase syntaxiquement correcte telle que :

"Je voudrais avoir le poste de Monsieur X"

le système est capable de reconnaître les phrases suivantes :

"J'aimerais avoir le poste de Monsieur X"

"Je veux le poste de Monsieur X"

"Je voudrais avoir le poste de X"

"Je désirerais avoir, euh !, le poste de, pardon, de Mr X"

"Je voudrais avoir, si c'est possible, le poste de X"
etc...

La grammaire choisie détermine donc un noyau de référence permettant de guider la reconnaissance d'un ensemble plus vaste plutôt que de définir entièrement les phrases acceptées.

1-2-2- Le lexique :

Pour la reconnaissance de cet ensemble de phrases le système n'utilise que le lexique correspondant au noyau de référence : soit 18 terminaux vrais et 2 terminaux de type

Figure C-1-2
Grammaire du langage

< PHRASE >	::=	<u><APPEL ></u> <DEMANDE> <DEMANDE>
< APPEL >	::=	<u>Allo</u> <SALUT> <u>allo</u> <SALUT>
< SALUT >	::=	<u>Bonjour</u> <TYPE> < <u>Bonjour</u> > <TYPE>
< DEMANDE >	::=	<REQUETE> <POLI> <REQUETE> <POLI>
< POLI >	::=	<u>s'il vous plaît</u> <u>Merci</u> <u>s'il vous plaît</u> <u>Merci</u>
< REQUETE >	::=	<INT> <AFFIR >
< INT >	::=	<u>Pourrais-je</u> <DEM > <u>est-ce que je pourrais</u> <DEM >
< AFFIR >	::=	<u>Je voudrais</u> <DEM > <u>passer-moi</u> <OBJET> <OBJET >
< DEM >	::=	<u>Parler à</u> <PERS > <u>avoir</u> <OBJET >
< PERS >	::=	<TYPE > <NOM >
< OBJET >	::=	<u>le</u> <QUALIF > <u>de</u> <PERS > <PERS > <u>le</u> <NUM > <u>le</u> <QUAL 1 > <NUM >
< QUALIF >	::=	<u>Bureau</u> <QUAL 1 >
< QUAL 1 >	::=	<u>Poste</u> <u>numéro</u>
< TYPE >	::=	<u>Monsieur</u> / <u>Madame</u> / <u>Mademoiselle</u>
< NOM >	::=	<u>Albert</u> / <u>Dupont</u> / <u>Durand</u> / <u>Pierre</u> / <u>Pierrel</u> /
< NUM >	::=	<u>221</u> / <u>222</u> / <u>223</u> / <u>239</u> / <u>240</u> / <u>241</u>

lexique. Ces derniers regroupent les noms de personne et les numéros de poste téléphonique inscrits dans l'annuaire du standard. Le vocabulaire ainsi utilisé comporte 31 mots ou groupes de mots donnés dans le tableau ci-dessous. (La codification de ces différents mots est donnée entre parenthèses).

a) les terminaux vrais de la grammaire noyau

(1) ALLO	(10) AVOIR
(2) BONJOUR	(11) LE
(3) S'IL VOUS PLAIT	(12) DE
(4) MERCI	(13) BUREAU
(5) POURRAIS-JE	(14) POSTE
(6) EST-CE-QUE-JE POURRAIS	(15) NUMERO
(7) JE VOUDRAIS	(16) MONSIEUR
(8) PASSEZ-MOI	(17) MADAME
(9) PARLER A	(18) MADEMOISELLE

b) les terminaux de type lexique de la grammaire noyau

(19) NOM	(20) NUMERO
ALBERT (101)	221 (201)
DUPONT (102)	222 (202)
PIERREL (103)	223 (203)
BOUCHET (104)	339 (204)
PIERRE (105)	340 (205)
DURAND (106)	341 (206)

c) les mots utilisés dans le dialogue dont nous verrons l'utilité lors de l'étude de la procédure de dialogue (chapitre 2).

On peut répartir ces mots en deux groupes :

- les mots communs au dialogue et au vocabulaire de la grammaire noyau :

POSTE

MONSIEUR

MADAME

MADEMOISELLE

et les terminaux de type lexique NOM et NUMERO

- les mots propres au dialogue.

Pour notre application, cet élément est réduit au minimum, il ne comporte qu'un élément : OUI.

Tous ces mots correspondent à des parties de réponses possibles aux questions posées par la machine pour lever les ambiguïtés ou récupérer les erreurs après une première reconnaissance.

1-3- INFORMATIONS SEMANTIQUES :

Dans le cadre de cette étude sur la reconnaissance automatique de la parole, nous avons opté pour une définition très large de la sémantique : seront appelées informations sémantiques toutes les informations ne relevant pas strictement de la syntaxe définie, pour le langage noyau, par la grammaire donnée en paramètre du système. Ce type d'information est nécessaire à 2 niveaux : d'une part pour la reconnaissance faite par le système exposé en partie B, d'autre part pour la procédure de dialogue qui doit permettre de déclencher l'action correspondant à la demande orale reconnue.

Le système de reconnaissance de phrase est guidé par la syntaxe du langage noyau : la signification des mots utilisés et des phrases générées par la grammaire donnée en figure C-1-2 est non ambiguë ; nous n'explicitons donc pas ici la sémantique de ce langage. Néanmoins, comme nous l'avons vu lors de l'étude de l'analyseur lexicographique, il nous faut une liste de mots caractéristiques au niveau sémantique (cf paragraphe B-2-2-3- et B-4-2-). Cette liste doit contenir les mots sans lesquels un message oral, émis dans le contexte de cette application, ne peut être compris : dans le cadre de ce langage de communication avec un standard téléphonique, nous avons limité cet ensemble aux deux lexiques NOM et NUMERO.

De plus, dans la procédure de dialogue, l'action correspondant à la demande reconnue par le niveau syntaxique sera provoquée, compte tenu du contexte propre à l'application. Cette action finale du système peut être ici de trois types : branchement d'une ligne, mise en attente d'un correspondant, fin d'une communication. Pour ce faire, il faut diverses informations sémantiques supplémentaires dont une table de correspondance entre noms et numéros et une liste, sans cesse tenue à jour, des lignes occupées.

1-4- VALIDITE DU LANGAGE :

Le langage défini par la grammaire C-1-2- a été construit, comme il a déjà été dit, à partir des résultats fournis par un pointage fait "en espionnant" un standard téléphonique. Le nombre d'appels consécutifs provenant de l'extérieur pris en compte est de 108.*

Si on ajoute à la grammaire la possibilité d'appeler des "services" : secrétariats, laboratoires, services administratifs (cette extension sera prochainement implémentée) et compte tenu de la possibilité d'insertion de mots parasites, de substitution ou d'élosion d'un terminal du langage :

97 appels (soit 90 %) sont admis par ce langage
11 appels (soit 10 %) ne sont pas admis.

Il est à noter que la quasi-totalité des appels non admis correspondent à des demandes de renseignements assez complexes auxquelles la standardiste répond le plus souvent en branchant le correspondant sur un secrétariat ou un service de renseignement approprié. Dans le cadre d'un système automatisé, si le correspondant est prévenu, on peut certainement penser que ces demandes se traduiraient par un appel à un service de renseignements ou à un secrétariat, sinon le système lui-même effectue ce branchement lors de la phase de traitement des erreurs de reconnaissance.

* Nous remercions M. Claude ROCHE, du Laboratoire Central de l'Armement pour nous avoir fourni la transcription de la bande magnétique "d'espionnage".

C H A P I T R E 2

LA PROCEDURE DE DIALOGUE PROPRE A L'APPLICATION

2-1- BUT A ATTEINDRE :

Dans le cadre de notre application d'automatisation d'un standard téléphonique, le but à atteindre est de rendre le meilleur service possible à l'utilisateur du standard : ceci correspond soit à un branchement de la ligne demandée si celle-ci est libre, soit dans le cas contraire à une mise en attente ou une fin de communication suivant le désir de l'usager.

Les qualités d'un tel service doivent être avant tout la rapidité, l'efficacité, et la fiabilité ; elles déterminent en grandes parties les caractéristiques du traitement à effectuer à partir des résultats de la reconnaissance :

- la rapidité nécessite un traitement simple dont le temps de réponse soit peu important en vue d'un traitement proche du temps réel. Mais cet aspect est très relatif et dépend beaucoup du matériel utilisé.

- l'efficacité correspond à un système capable de récupérer facilement les erreurs et dont la phase de dialogue, réduite au maximum ne provoque jamais de répétitions inutiles de la part du locuteur.

- la fiabilité enfin oblige le système à ne provoquer l'action correspondant à une demande que s'il est certain du résultat de la reconnaissance : toutes les ambiguïtés devront être levées préalablement et, sinon, il sera préférable d'avoir recours à un opérateur humain plutôt que de provoquer le branchement d'une ligne non désirée.

La mise en oeuvre de ces trois caractéristiques principales se fait grâce à une procédure de dialogue qui a la triple tâche de vérifier la reconnaissance, lever les ambiguïtés ou récupérer les erreurs, puis seulement de provoquer le branchement correspondant à la communication demandée. Dans notre cas, les réponses du système sont pour l'instant imprimées, mais il est évident qu'en pratique ce dialogue homme-machine nécessite l'emploi d'un système de synthèse vocale.

2-2-LES DIFFERENTS CAS A CONSIDERER :

2-2-1- Phrase bien reconnue :

Ce premier cas correspond à une phrase dont la reconnaissance est certaine au sens où nous l'avons défini au paragraphe B-6-1-. Ceci nécessite trois conditions :

- un taux de réussite SC supérieur à un seuil donné.
- la syntaxe de la phrase reconnue correspond rigoureusement à la grammaire donnée à la figure C-1-2-.
- la pile des hypothèses validées par le niveau lexicographique et non traitée par le niveau syntaxique ne contient pas de mots caractéristiques au niveau sémantique, NOM ou NUMERO.

Le système teste alors si la ligne demandée est libre ou occupée. Si la ligne est libre, il branche la communication, sinon il demande au correspondant s'il désire attendre ou non et, suivant la réponse de celui-ci, provoque soit la mise en attente soit la fin de la communication.

2-2-2- Phrase reconnue avec ambiguïtés :

Le second cas correspond à une phrase dont la reconnaissance fait apparaître une ou plusieurs causes d'ambiguïtés qui peuvent être :

- un taux de réussite SC trop faible,
- un recours, lors de la reconnaissance, aux hypothèses supplémentaires, émises par l'analyseur lexicographique, ce qui correspond à une liberté prise par rapport à la grammaire donnée à la figure C-1-2-.

- une reconnaissance multiple de terminaux correspondant aux mots caractéristiques au niveau sémantique :
NOM ou NUMERO.

Le système demande une confirmation par une question du type : "Vous avez bien demandé <objet de la demande> ?". Si la réponse est "oui", on se retrouve alors dans le cas d'une demande vérifiée, vu en C-2-2-1-. Sinon, le système attend que le correspondant renouvelle sa demande et, après traitement, on se retrouve soit dans le cas d'une phrase bien

reconnue soit dans le cas d'une erreur (si la reconnaissance est mauvaise).

2-2-3- Phrase non reconnue : récupération des erreurs :

Il reste enfin à considérer le cas correspondant aux échecs de la reconnaissance d'une phrase. Le système que nous avons implémenté utilise alors une procédure d'erreur qui admet deux niveaux d'erreurs successifs dans une même communication.

Le premier niveau provoque une demande de répétition de la phrase prononcée et réinitialise le processus de reconnaissance.

Le second niveau, quant à lui, intervient toujours après une première répétition de la demande et provoque le branchement à un opérateur humain.

Cette gestion à deux niveaux permet d'accepter la répétition de la phrase non reconnue sans pour autant courir le risque d'un bouclage sans fin lors du traitement d'une communication.

2-3- REALISATION DE LA PROCEDURE DE DIALOGUE :

2-3-1- Traitement des réponses en cours de dialogue :

Au cours de la phase de dialogue, le système doit reconnaître les réponses données par l'utilisateur, parmi lesquelles il distingue deux grandes classes, d'une part les répétitions complètes de la demande de départ - dans ce cas, le système se réinitialise et la reconnaissance se fait de la manière exposée en partie B-, d'autre part, les réponses particulières correspondant à une question posée par le système. Le traitement réalisé est alors spécifique, compte tenu de la question posée. Il consiste le plus souvent à reconnaître le mot "oui" en tête de la réponse. Dans les autres cas la reconnaissance est guidée par un petit automate fini qui permet de traiter des groupes de quelques mots. Il est initialisé au départ après une recherche dans la chaîne phonémique correspondant à la réponse d'un mot-clé. Ainsi, après la question "Vous avez bien demandé le poste 340 ?" si la réponse n'est pas

"oui" mais de type "Non le poste X", il y a recherche du groupe "le poste" dans la phrase et, seulement en cas de réussite, reconnaissance du numéro. On utilise donc dans cette phase de dialogue le module de reconnaissance de mots défini partie B chapitre 5 en liaison avec une recherche de mots clés.

Cet algorithme de recherche d'un mot clé peut se définir ainsi : trouver la représentation d'un mot X dans la chaîne phonémique représentant la réponse fournie par l'utilisateur. Pour cela, si nous appelons LPR la longueur phonémique de la réponse et LPM(X) la longueur phonémique minimale du mot X, nous effectuons des tentatives de reconnaissance de ce mot à partir des positions phonémiques variant de 1 à LPR-LPM(X) dans la chaîne représentant la réponse à traiter.

Une autre méthode pourrait consister à faire une reconnaissance du mot X à partir du spectre de la phrase. Mais, comme nous l'avons déjà dit, nous n'utilisons pas les informations spectrales pour l'instant.

2-3-2- Exemples de dialogue possible :

Plutôt que de présenter une analyse détaillée de ce module de dialogue, nous avons préféré regrouper ici quelques exemples de dialogue possible, accompagnés de commentaires et recouvrant les différents cas traités.

On notera S.A. pour Standard Automatisé
C. pour Correspondant.

S.A. : "Allo standard automatique de....., je vous écoute"

C. : "Allo je voudrais parler à Monsieur Dupont s'il vous plaît, merci."

(reconnaissance de la phrase)

S.A. : "Vous avez bien demandé Monsieur Dupont ?"

(la machine répète la demande de manière interrogative car il y a ambiguïté dans la reconnaissance de la phrase, par exemple possibilité de confondre Dupont et Durand).

C : "Oui"

(reconnaissance de "oui" : la demande précédente est validée).

S.A. : "Vous avez en ligne, Monsieur Dupont"

(la ligne étant libre le système effectue alors le branchement)

S.A. : "Allo standard automatique de...., je vous écoute"

C. : "Je voudrais madame Albert"

S.A. : "Je vous passe madame Albert"

(le système a reconnu la phrase sans ambiguïté possible, il s'apprête à effectuer le branchement.

Si le correspondant n'ajoute rien ou seulement oui ou merci, il effectue le branchement sinon il considère qu'il y a ambiguïté).

C. :

S.A. : Vous avez en ligne madame Albert

S.A. : "Allo standard automatique de..... je vous écoute"

C. : "Je voudrais le poste 222"

S.A. : "Vous avez bien demandé le poste 220 ?"

(le système pense avoir reconnu la phrase mais il y a ambiguïté sur la reconnaissance; en fait il s'agit ici d'une mauvaise interprétation)

C. : "Non le poste 222"

(la réponse n'étant pas de type "oui" il y a recherche du mot "le poste" puis reconnaissance du numéro)

S.A. : "le poste 222 est occupé ; voulez-vous patienter ?"

(la reconnaissance de la réponse précédente s'est bien effectuée mais le poste demandé est occupé)

C : "Non, Merci, je rappellerai"

(la réponse n'est pas de type oui, le système termine la communication).

S.A. : "Allo standard automatique de.... je vous écoute"

C. : "Bonjour je voudrais le poste 332"

- S.A. : "Je vous passe le poste 330"
 (la machine pense avoir reconnu correctement la phrase
 mais il y a erreur sur le numéro)
- C. : "Allo je veux le 332 pas le 330"
 (la réponse n'est pas du type "oui" "merci" ou silence
 il y a demande de précision)
- S.A. : "Vous avez bien demandé le poste 330 ?"
- C. : "Non le 332"
 (le système ne réussit pas à comprendre la demande)
- S.A. : "Désolé, je n'ai pas compris, veuillez répéter votre
 demande"
 (demande de répétition de la part du système et
 réinitialisation de la reconnaissance)
- C. : "Je voudrais le poste 332"
- S.A. : "Je vous passe le poste 332"
 (bonne reconnaissance sans ambiguïté)
- C. : "Oui c'est cela"
 (reconnaissance de "oui" en début de phrase)
- S.A. : "Vous avez en ligne le poste 332"
- S.A. : "Allo standard Automatique de...., je vous écoute"
- C. : "Euh..... Je voudrais un renseignement...."
 (le correspondant se trouve dérouté devant le standard
 automatisé car sa demande n'est pas très précise)
- S.A. : "Désolé, je n'ai pas compris, veuillez répéter, je vous
 écoute"
 (la machine ne reconnaît pas la phrase qui ne fait
 pas partie du langage. Elle provoque une erreur et
 demande de répéter la demande).
- C. : "Euh... savez-vous s'il y a possibilité de visiter les
 laboratoires ?"
 (le correspondant a précisé sa demande mais elle ne fait
 pas partie du langage accepté).
- S.A. : "Veuillez patienter, je vous passe la standardiste
 (la machine ne reconnaît pas la phrase, elle provoque
 une seconde erreur et passe le relai à un opérateur
 humain).

C H A P I T R E 3

RESULTATS OBTENUS EN TEMPS DIFFERE

3-1- EXEMPLE DE PHRASES RECONNUES :

Jusqu'à présent tous les tests ont été effectués en temps différé sur l'ordinateur CII 10070 puis IRIS 80 de l'IUCA de NANCY. Les données en entrée étaient formées de suites phonémiques tenant compte des performances actuelles du système acoustique. Voici à titre d'exemple quelques phrases correctement reconnues lors de ces tests :

"Allo je voudrais parler à Madame Durand"

"Je voudrais parler à Monsieur Pierre"

"Je voudrais parler à Madame Pierrel"

"Je voudrais le poste 339"

"Bonjour, je voudrais le poste 340, merci"

"Allo Bonjour Madame Je voudrais le poste 323 s'il vous plait, merci"

"Je voudrais parler à, Euh, à Madame Albert, Merci"

"Je voudrais avoir poste 223"

"Bonjour je désirerais parler à Madame Bouchet"

"Allo, Mademoiselle, est ce que je peux avoir Madame Dupont"

La figure C-3-1- donne quelques exemples de chaînes de phonèmes utilisées en entrée lors de ces tests et permet, ainsi, d'avoir un aperçu de leur qualité.

3-2- EXEMPLES DE RESULTATS DU SYSTEME COMPLET :

3-2-1- Présentation et discussion de résultats du système complet :

Le tableau C-3-2- reproduit un listing obtenu au cours des tests du système sur l'IRIS 80 de l'IUCA. Sur ce listing, les phrases précédées de " >>> COR" correspondent au texte fourni à la machine en même temps que la donnée représentant la chaîne phonémique correspondante (ce texte ne nous sert qu'à présenter des résultats plus clairs ; il n'intervient en aucune manière dans la reconnaissance). Les autres phrases

a) z a v u b z e a v w m ə s ə b ʏ p ô
 v ə z ʏ d e e e z a b e s e d u k â
 z e z g e i ə z d a f i g o t o

"Je voudrais avoir Monsieur Dupont"

b) z ə v u u b i e a v a z ə p o s k ə p w f â k a â t ə
 z a z ʏ d e e e z w l a t u s t a t a s ô p ə ê k a
 v e z o o g e i ə z o e e k f p e k o s t e ô p â

"Je voudrais avoir le poste 340"

c) z u d e l p ə s ə p w s â t â n ə f
 v ʏ b i z t o s a t o s ô p ô m ʏ s
 z o g e e k a e k f ê k ê d s

"J'voudrais l'post' 339"

d) a l o z a z ʏ u b i e p ə a e b a ə d a â
 e z u v ə v u e d e e t a e o d ə o g o ô
 ə ə z e z o g e i k e a g o a b ə m

Allo je voudrais parler à madame

d ʏ o m e l s e
 b u a b a z f i
 g a d ə s o

durand merci

Exemples de chaîne de phonèmes en entrée.

Figure C-3-1-

indiquent les réponses fournies par le système.

Ces tests ont été construits pas à pas, car on ne pouvait pas prévoir à priori les réponses du système qui ne travaille pas encore sous forme conversationnelle.

Les résultats de la simulation de ces dix communications téléphoniques différentes sont les suivants :

a) communication 1 : dans un premier temps la phrase du correspondant n'a pas été reconnue et le système a provoqué sa répétition. Cette seconde demande a été reconnue, mais il restait une ambiguïté qui est levée par la question suivante et la réponse "oui" fournie par le correspondant. La communication se termine sans mise en attente car la réponse à la question "voulez-vous patienter ?" ne commence pas par "oui".

b) communication 2 : c'est un exemple de phrase reconnue sans ambiguïté. La ligne étant libre, le branchement s'effectue tout de suite. On peut remarquer que la communication se termine par deux messages du système. Si, après le premier message : "Je vous passe Mme Bouchet", le correspondant était intervenu par une phrase ne commençant ni par "oui" ni par "merci", le système aurait demandé une précision sur la demande formulée.

c) communication 3 : Dans cette exemple on peut noter une mauvaise reconnaissance du numéro de poste : 339 au lieu de 340. La reconnaissance de "339" étant ambiguë, le système a demandé une précision au correspondant ; la réponse obtenue n'étant pas de type "oui", il a recherché dans cette réponse le mot clé "le poste" ce qui lui a permis de reconnaître correctement ensuite le numéro demandé : 340 et de provoquer ainsi le branchement correspondant.

d) communication 4 : cette communication donne un exemple de mauvaise reconnaissance pour des phrases faisant pourtant partie du langage traité. Cet exemple correspond à une mauvaise prononciation de "Durand"; à la place du "r"

une plosive a été détectée forçant ainsi la reconnaissance du "Dupont". On voit ainsi les deux niveaux d'erreurs utilisés par le système : dans un premier temps il demande une répétition de la demande et ensuite il passe le relais à un standardiste humaine.

e) communication 5 : à la question du système : "vous avez bien demandé le poste 223 ?" le correspondant a répondu par "oui". La détection de cette réponse sert de confirmation au système qui provoque donc ensuite le branchement au bon numéro.

f) communication 6 : elle est du même type que la communication 3 mais l'ambiguïté porte ici sur un nom et non plus sur un numéro. De plus, la ligne étant occupée, le système propose une mise en attente refusée par le correspondant.

g) communication 7 : On peut noter dans cet exemple que la question formulée par le système pour lever l'ambiguïté sur le nom et la réponse "oui" donnée par le correspondant ont permis de confirmer la reconnaissance de la première demande formulée. De plus, la réponse "oui" à la proposition de mise en attente provoque bien cette mise en attente.

h) communication 8 : deux phrases sont à noter dans cet exemple: tout d'abord une erreur de numéro dans la première demande formulée par le correspondant se trouve rattrapée par le dialogue ultérieur : de 223 le correspondant passe à 221, correspondant au numéro véritablement souhaité. De plus, dans la première demande, une partie de la chaîne phonémique de départ ne correspondait à rien et malgré cela, le système a utilisé le reste de la phrase pour mener à bien la reconnaissance.

i) communication 9 : elle correspond à une reconnaissance parfaite de même type que la communication 2.

j) communication 10 : Cette communication correspond à des demandes ne faisant pas partie du langage accepté. On peut y noter les deux niveaux d'erreurs : le premier conduit à une demande de répétition de la phrase alors que le second provoque un recours à un opérateur humain.

DEBUT DE SESSION

a)

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS POSTE 339'

DESOLE, JE N AI PAS COMPRIS. VEUILLEZ REPETER, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS LE POSTE 339'

VOUS AVEZ BIEN DEMANDE LE POSTE 339?

>>> COR: 'OUI'

LE POSTE 339 EST OCCUPE. VOULEZ-VOUS PATIENTER?

>>> COR: 'NON MERCI, JE RAPELLERAI.'

FIN DE COMMUNICATION

b)

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS PARLER A MADA ME BOUCHET'

JE VOUS PASSE MME. BOUCHET

VOUS AVEZ EN LIGNE MME. BOUCHET

FIN DE COMMUNICATION

c)

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS AVOIR LE POSTE 340'

VOUS AVEZ BIEN DEMANDE LE POSTE 339?

>>> COR: 'NON LE POSTE 340'

VOUS AVEZ EN LIGNE LE POSTE 340

FIN DE COMMUNICATION

d/

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS AVOIR MADAME DURAND '

DESOLE, JE N AI PAS COMPRIS. VEUILLEZ REPETER, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS AVOIR MADAME DURAND '

JE VOUS PASSE MME. DUPONT

>>> COR: 'ALLO JE VOUDRAIS PARLER A MADAME DURAND '

VOUS AVEZ BIEN DEMANDE MME. DUPONT ?

>>> COR: 'JE VOUDRAIS PARLER A MADAME DURAND '

VEUILLEZ PATIENTER JE VOUS PASSE LA STANDARDISTE

FIN DE COMMUNICATION

e/

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'ALLO BONJOUR MADAME JE VOUDRAIS LE POSTE 223 S.V.P. MERCI'

VOUS AVEZ BIEN DEMANDE LE POSTE 223?

>>> COR: 'OUI'

VOUS AVEZ EN LIGNE LE POSTE 223

FIN DE COMMUNICATION

f/

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS PARLER A MONSIEUR PIERREL'

VOUS AVEZ BIEN DEMANDE MR. PIERRE ?

>>> COR: 'NON, MONSIEUR PIERREL'

LE POSTE DE MR. PIERREL EST OCCUPE. VOULEZ-VOUS PATIENTER?

Tableau C.3.2.(b)

>>> COR: 'NON MERCI, JE RAPELLERAI.'

FIN DE COMMUNICATION

81

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS PARLER A MONSIEUR PIERREL'

VOUS AVEZ BIEN DEMANDE MR. PIERREL ?

>>> COR: 'OUI'

LE POSTE DE MR. PIERREL EST OCCUPE. VOULEZ-VOUS PATIENTER?

>>> COR: 'OUI'

LIGNE EN ATTENTE

FIN DE COMMUNICATION

R1

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'ALLO BONJOUR ?????? JE VOUDRAIS LE POSTE 223 S.V.P. MERCI'

VOUS AVEZ BIEN DEMANDE LE POSTE 223?

>>> COR: 'JE VOUDRAIS AVOIR LE POSTE 221'

VOUS AVEZ EN LIGNE LE POSTE 221

FIN DE COMMUNICATION

i

DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'JE VOUDRAIS PARLER AMADAME ALBERT'

JE VOUS PASSE MME. ALBERT

VOUS AVEZ EN LIGNE MME. ALBERT

FIN DE COMMUNICATION

Tableau C.3.2.(c)



DEBUT DE COMMUNICATION

STANDARD AUTOMATIQUE DE ****, JE VOUS ECOUTE

>>> COR: 'ALLO JE VOUDRAIS UN RENSEIGNEMENT'

DESOLE. JE N AI PAS COMPRIS. VEUILLEZ REPETER, JE VOUS ECOUTE

>>> COR: 'EUH CE SERAIT POUR UN RENSEIGNEMENT'

VEUILLEZ PATIENTER JE VOUS PASSE LA STANDARDISTE

FIN DE COMMUNICATION

FIN DE SESSION

JMP

STOP 0

JOB STEP 03 TERMINATED AT 10*44*20* AFTER 0000.07 MIN

• CORE USED 0032 DISC USED 0014 WAIT TIME 0000.03

TIME	TIME*CORE	CORE-USE	TIME*DISC	DISC-USE
00.66	29.12	92%	19.85	41%

DS	DS	SHR	SHR	CR	CP
I/O-BYTES	I/O-CALLS	I/O-BYTES	I/O-CALLS	CARDS	CARDS
1610528	838	118784	116	552	00
K 1	05				
	01				
	01				
	00.59				
	00				
	00				

SELFJOB TERMINATED PIERREL

IUCA 10*44*21*4

Tableau C.3.2.(d)

3-2-2- Exemples de résultats intermédiaires :

Les tableaux C-3-3- et C-3-4- donne des exemples commentés de résultats intermédiaires obtenus lors de la reconnaissance de deux phrases différentes.

Dans le tableau C-3-3- sont regroupés les résultats intermédiaires relatifs à la reconnaissance de la phrase "Allô je voudrais parler à madame Durand, merci" dont la chaîne phonémique correspondante est donnée en d) figure C-3-1-. Ces résultats correspondent à des ordres "ØUTPUT" fortran placés après la vérification des hypothèses.

NTER indique le numéro du terminal reconnu (cf
paragraphe C-1-2-2-)

J est l'indice de positionnement de la grammaire

IR indique le nombre de mots reconnus dans le
contexte gauche déjà traité

P est le pointeur courant dans la chaîne phonémique
à reconnaître

H et HF sont des variables utilisées lors de la sauvegarde
du contexte de l'Analyse syntaxique.

On peut remarquer que cette phrase correspond strictement à la grammaire du langage noyau donnée en figure C-1-2- et que la reconnaissance s'est faite sans aucun retour en arrière.

Le tableau C-3-4- correspond quant à lui à la reconnaissance de la phrase "J'voudrais l'post' 339" dont la chaîne phonémique correspondante est donnée en c) figure C-3-2-. On retrouve le même "OUT PUT" 'Terminal Trouvé' accompagné de divers autres :

a) PROB : après toute reconnaissance réussie cet
"OUT PUT" donne la valeur du taux de réussite
de cette reconnaissance (RESULT)

b) HYPOTHESE : donne la liste des hypothèses supplé-
mentaires émises par l'analyseur lexicographique
(cf partie B paragraphe 4-2-)

TERMINAL TROUVE NTER = 1 J = 1 H = -1 IR = 0 HF = -1 P = 4	reconnaissance de "Allo" (phonèmes 1 à 4)
TERMINAL TROUVE NTER = 7 J = 2 H = -1 IR = 1 HF = 1 P = 11	reconnaissance de "Je voudrais" (phonèmes 4 à 11)
TERMINAL TROUVE NTER = 9 J = 27 H = 7 IR = 2 HF = 7 P = 16	reconnaissance de "Parler à" (phonèmes 11 à 16)
TERMINAL TROUVE NTER = 17 J = 33 H = 9 IR = 3 HF = 9 P = 21	reconnaissance de "Madame" (phonèmes 16 à 21)
TERMINAL TROUVE NTER = 106 J = 39 H = 11 IR = 4 HF = 13 P = 25	reconnaissance de "Durand" (phonèmes 21 à 25)
TERMINAL TROUVE NTER = 4 J = 11 H = 3 IR = 5 HF = 13 P = 30	reconnaissance de "Merci" (phonèmes 25 à 30)

PHRASE RECONNUE AVEC UNE PROBABILITE EGALE A 1.00

1 7 9 17 106 4

Allo, je voudrais parler à madame Durand, Merci.

Tableau C-3-3-

```

PROP
J = 1
RESULT(J,2) = .700000
TERMINAL TROUVE
NTER = 7
J = 2
H = -1
IR = 0
HF = -1
P = 5
HYPOTHESE
HYPO
TAB(I,1) = 9
TAB(I,1) = 10
TAB(I,1) = 16
TAB(I,1) = 17
TAB(I,1) = 18
TAB(I,1) = 11
TAB(I,1) = 16
TAB(I,1) = 17
TAB(I,1) = 18
SELECN
SELECN
SELECN
SELECN
SELECN
PROR
J = 1
RESULT(J,2) = 1.00000
TERMINAL TROUVE
NTER = 11
J = 36
H = 9
IR = 1
HF = 9
P = 6
PROR
J = 1
RESULT(J,2) = .933333
PROB
J = 1
RESULT(J,2) = .933333
TERMINAL TROUVE
NTER = 14
J = 42
H = 11
IR = 2
HF = 11
P = 9
HYPOTHESE
HYPO
TAB(I,1) = 12
TAB(I,1) = 16
TAB(I,1) = 17
TAB(I,1) = 18
SELECN
PROB
J = 1
RESULT(J,2) = 1.00000
SELECN
SELECN

```

reconnaissance de "Je voudrais"
avec un score de réussite de 0,7.
(phonèmes de 1 à 5)

hypothèses émises par l'analyseur
lexicographique après l'échec de la
reconnaissance de "Avoir" et "parler à"
(avoir, parler à, Monsieur (2X),
Madame (2X), Mademoiselle (2X), le)

tentatives de reconnaissance des
hypothèses

reconnaissance du terminal "le", score
de réussite de 1
(phonèmes 5,6)

reconnaissance du terminal "poste"
score de réussite de 0.933
(phonèmes 6 à 9)
ce terminal est noté deux fois
a) dans le cas de "poste de M.X"
b) dans le cas de "poste < Numéro > "

dans le premier cas, émission d'hypothèses
après mauvaise reconnaissance de terminal
"de" (de, Monsieur, Madame, Mademoiselle)

tentative de reconnaissance des hypothèses
réussite pour le terminal "de" après
une hypothèse d'insertion d'un mot
parasite

```

SELECN
TERMINAL TROUVE
NTER = 12
J = 43
H = 11
IR = 3
HF = 15
P = 18
TERMINAL TROUVE
NTER = 14
J = 50
H = 11
IR = 2
HF = 15
P = 9
PROR
J = 1
RESULT(J,2) = .755556
PROR
J = 2
RESULT(J,2) = .755556
PROR
J = 3
RESULT(J,2) = .866667
PROR
J = 4
RESULT(J,2) = .755556
PROR
J = 5
RESULT(J,2) = .866667
PROR
J = 6
RESULT(J,2) = .866667
PROR
J = 7
RESULT(J,2) = .977778
TERMINAL TROUVE
NTER = 204
J = 51
H = 11
IR = 3
HF = 17
P = 18

```

hypothèse validée : "de"
pour une reconnaissance se terminant
au phonème 18

après une tentative de poursuite de la
reconnaissance
le système opère un retour arrière et
considère le second cas possible avec
poste ("le poste < NUM > ")

reconnaissance du terminal de type
lexique.

7 chemins différents fournissent un
résultat.

Celui de taux le plus fort (0.977)
correspond à la réalisation notée 204
c.a.d. au numéro 339

fin de la reconnaissance

PHRASE RECONNUE AVEC UNE PROBABILITE EGALE A 1.00

7 11 14 204

"Je voudrais le poste 339"

Tableau C-3-4- (suite)

- c) SELEC indique un essai de sélection d'une chaîne phonémique représentant une des hypothèses générées par l'analyseur lexicographique dans la phrase à reconnaître.

On peut remarquer dans cet exemple que la phrase à reconnaître ne correspond pas strictement à la grammaire du langage noyau : le système a donc dû émettre une hypothèse d'élimination sur les mots "avoir" et "parler.à". On peut y noter également un essai infructueux de reconnaissance du groupe "le poste de M. X" qui a ensuite provoqué un retour arrière pour mener à bien la reconnaissance finale.

3-3- PERFORMANCES DU SYSTEME :

Il est encore impossible de donner les performances exactes du système mis en oeuvre. Seule, en effet, une implémentation en temps réel du système complet, niveau acoustique et niveau syntaxico-sémantique réunis, et une série de tests systématiques permettront de chiffrer exactement le temps de réponse et d'obtenir une valeur approchée du pourcentage de reconnaissance. Ceci n'a pas encore pu être réalisé, compte tenu des faibles moyens dont nous disposons actuellement.

Néanmoins, on peut dès maintenant noter des résultats intéressants tant au niveau temps de calcul qu'en possibilité de traiter des chaînes phonémiques assez fortement entachées d'erreurs.

Les phrases données en exemple au paragraphe C-3-1- sont correctement reconnues après un temps moyen inférieur à une demi-seconde et les exemples de dialogue fournis dans le tableau C-3-2- nous montrent que le système est capable de traiter 10 communications différentes en 5 secondes, soit environ une demi-seconde par communication. Si on se réfère aux résultats obtenus par le système acoustique (1), (43), qui

est capable de travailler en temps réel, on peut donc affirmer que le système complet peut le faire lui aussi dans le cadre de langage spécialisé comme celui traité ici. En effet, un temps de réponse inférieur à une seconde dans un dialogue peut être considéré à juste titre comme le résultat d'un système travaillant en temps réel. On peut noter cependant à nouveau le caractère relatif de la notion de temps réel : les résultats qui seront obtenus varieront énormément selon que l'implémentation du système se fera sur un ordinateur IRIS 80 ou Mitra 15, ou encore à l'aide d'une batterie de microprocesseur.

Ces résultats ne sont pas surprenants, car l'algorithme utilisé permet au module de reconnaissance analytique de ne traiter, à chaque instant, que le minimum de mots, compte tenu du contexte gauche de la phrase déjà reconnue. Les résultats intermédiaires, obtenus lors de la reconnaissance, montrent que le taux de retours en arrière effectués lors de l'analyse syntaxique est assez faible (de l'ordre de 10/100) et qu'en général seul le dernier mot reconnu est remis en cause. Les résultats ainsi obtenus pour cette application particulière sont très encourageants et semblent valider l'approche syntaxique que nous avons choisie.

C H A P I T R E I V

POURSUITE DU TRAVAIL

4-1- IMPLEMENTATION DU SYSTEME COMPLET EN TEMPS REEL :

4-1-1- Enchaînement avec le niveau acoustique :

Une poursuite directe de ce travail consiste à implémenter le système complet et à obtenir ainsi une chaîne de traitement continue à partir du micro, organe de saisie de données, jusqu'à la procédure de dialogue qui provoque l'action correspondant à la demande. Il s'agit donc de prévoir l'enchaînement du système acoustique avec le système syntaxico-sémantique. Cette jonction peut se faire de diverses manières : la plus simple consiste à considérer le résultat du système acoustique comme donnée du système syntaxico-sémantique, la jonction ne se faisant alors que dans un sens. Dans ce cas, la reconnaissance des phonèmes est faite indépendamment des contraintes syntaxiques et sémantiques. En fait la solution idéale est d'obtenir deux systèmes travaillant en étroite collaboration l'un avec l'autre : il serait intéressant de prévoir des interactions possibles du niveau syntaxique sur le niveau acoustique (66) spécialement en cas de mauvaise segmentation lors de l'obtention de la chaîne phonémique représentant la phrase.

Dans la solution adoptée jusqu'alors pour notre système qui correspond au cas simple, l'échange d'information entre ces deux phases de reconnaissance est très mauvais. L'utilisation de la chaîne phonémique seule provoque une perte considérable d'informations par rapport au signal de départ. Pour y remédier, il serait sans doute intéressant d'avoir recours à d'autres types d'informations parmi lesquels on peut noter :

- le spectre de la phrase (cf paragraphe B-2-1-2-) : c'est à partir du spectre que la chaîne phonémique est déterminée par le système acoustique après une compression importante d'informations qui peut entraîner certaines erreurs. La représentation spectrale est donc plus fidèle que la représentation phonémique, et il semble souhaitable de l'utiliser : on envisage ainsi un recours à une reconnaissance spectrale

par programmation dynamique lorsque des ambiguïtés subsistent après la phase de reconnaissance phonémique ou pour une recherche directe de mots clés dans le cas d'une approche globale.

- les indices de prosodie : nous ne tenons compte, dans notre système, d'aucun indice prosodique. C'est là pourtant une source fondamentale de connaissance pour la compréhension du discours continu dont l'étude approfondie peut apporter beaucoup d'amélioration sur la qualité de la reconnaissance. A titre d'exemple, nous avons pensé utiliser un détecteur de mélodie (Melographe CNET/ETA) pour tester si une phrase est interrogative ou affirmative. Des contraintes matérielles nous ont empêché de mettre en place cette application. Il y a pourtant là une source non négligeable de données qui doit permettre une simplification de la grammaire du langage traité et ainsi accélérer le processus d'analyse syntaxique utilisé.

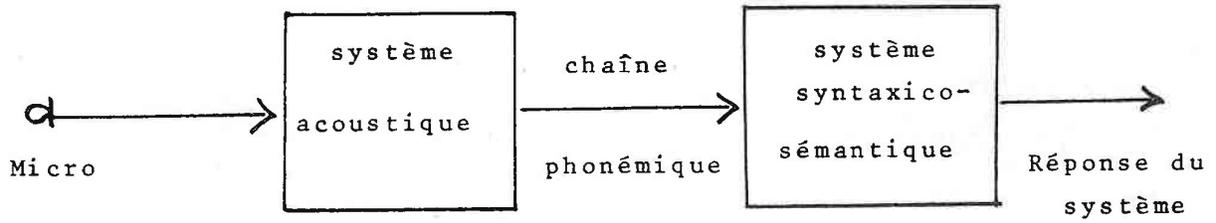
- les contraintes linguistiques : non utilisées dans le système actuel, ces contraintes peuvent être très utiles dans la détermination des frontières de mots, ce qui simplifierait la sélection d'un mot dans la chaîne phonémique. Une étude statistique présentée par J.P. HATON (1) fournit à ce sujet des renseignements intéressants sur les commencements et les terminaisons de mots en français.

Ainsi, l'enchaînement des niveaux acoustique et syntaxico-sémantique n'est pas un problème simple, si du moins on cherche à utiliser le maximum d'informations fournies dans l'énoncé oral de départ.

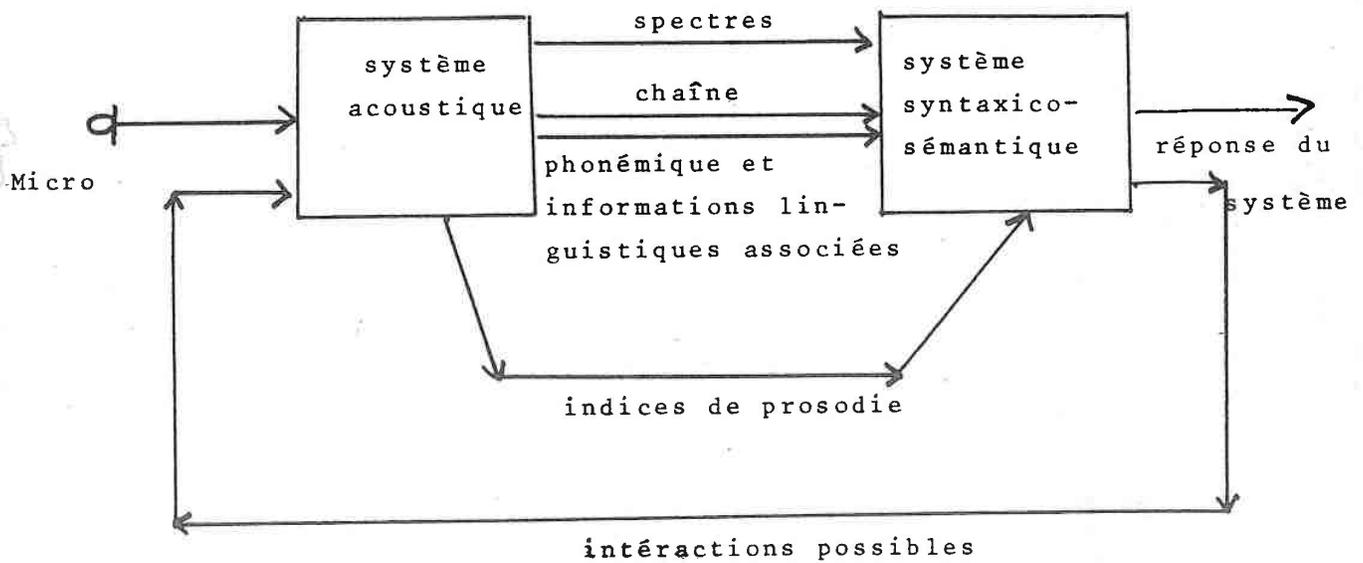
La figure C-4-1- schématise les jonctions possibles dans le cas le plus simple et dans le cas idéal.

4-1-2- Système temps réel travaillant en mode conversationnel :

L'implémentation complète du système de reconnaissance doit se faire en ligne et en mode conversationnel afin de permettre un dialogue entre l'utilisateur et la machine. Nous espérons pouvoir réaliser un tel système à Nancy dans un avenir assez proche. On peut le décomposer en quatre grandes parties.



Jonction simple



Jonction idéale

Jonction entre les niveaux acoustique et syntaxico-sémantique

Figure C-4-1-

a) un poste de saisie des données :

il est très simple dans le cas de la parole et se décompose uniquement d'un micro haute fidélité.

b) une partie reconnaissance :

elle regroupe les niveaux acoustique et syntaxico-sémantique, et peut être implémentée sur un mini-calculateur -dans notre cas un CII Mitra 15- possédant une mémoire secondaire de type disque magnétique.

c) une partie traitement de la demande reconnue :

Elle dépend énormément de l'application mise en oeuvre : gestion de banque de données dans le cas d'une consultation orale de fichier, programmes de service dans le cas de commande orale d'ordinateurs etc.... Cette partie peut être nettement dissociée de la partie reconnaissance et, à l'image des réalisations en cours dans le projet ARPA aux Etats-Unis, on peut penser que ce traitement doit se faire sur un autre calculateur plus important, le mini-calculateur orienté vers la reconnaissance de la parole étant alors considéré comme un périphérique. Nous envisageons, dans ce sens, d'utiliser la connexion MITRA 15-IRIS 80 (de l'IUCA). Dans un tel cas, la limite entre le travail effectué par l'un ou par l'autre peut varier. Il est en effet possible de réduire la charge du mini-calculateur en reportant le traitement syntaxico-sémantique au niveau de l'ordinateur central. Cette configuration correspond aux nombreux cas pratiques où un terminal "intelligent" interroge un ordinateur central : réservations, gestion de stocks, renseignements, etc...

d) un poste de sortie de résultats : l'idéal est de posséder un synthétiseur de la parole. Mais on peut aussi utiliser des sorties écrites (imprimante) ou visuelles (console de visualisation).

De plus, on peut envisager, pour l'analyse et la reconnaissance, la réalisation de processus cablés spécialisés permettant d'alléger le traitement fait par programme : l'analyse spectrale du signal vocal peut ainsi se faire par un étage hardware, de même que la détection de mélodie ou diverses autres fonctions.

La figure C-4-2- donne un schéma général de l'implantation d'un tel système.

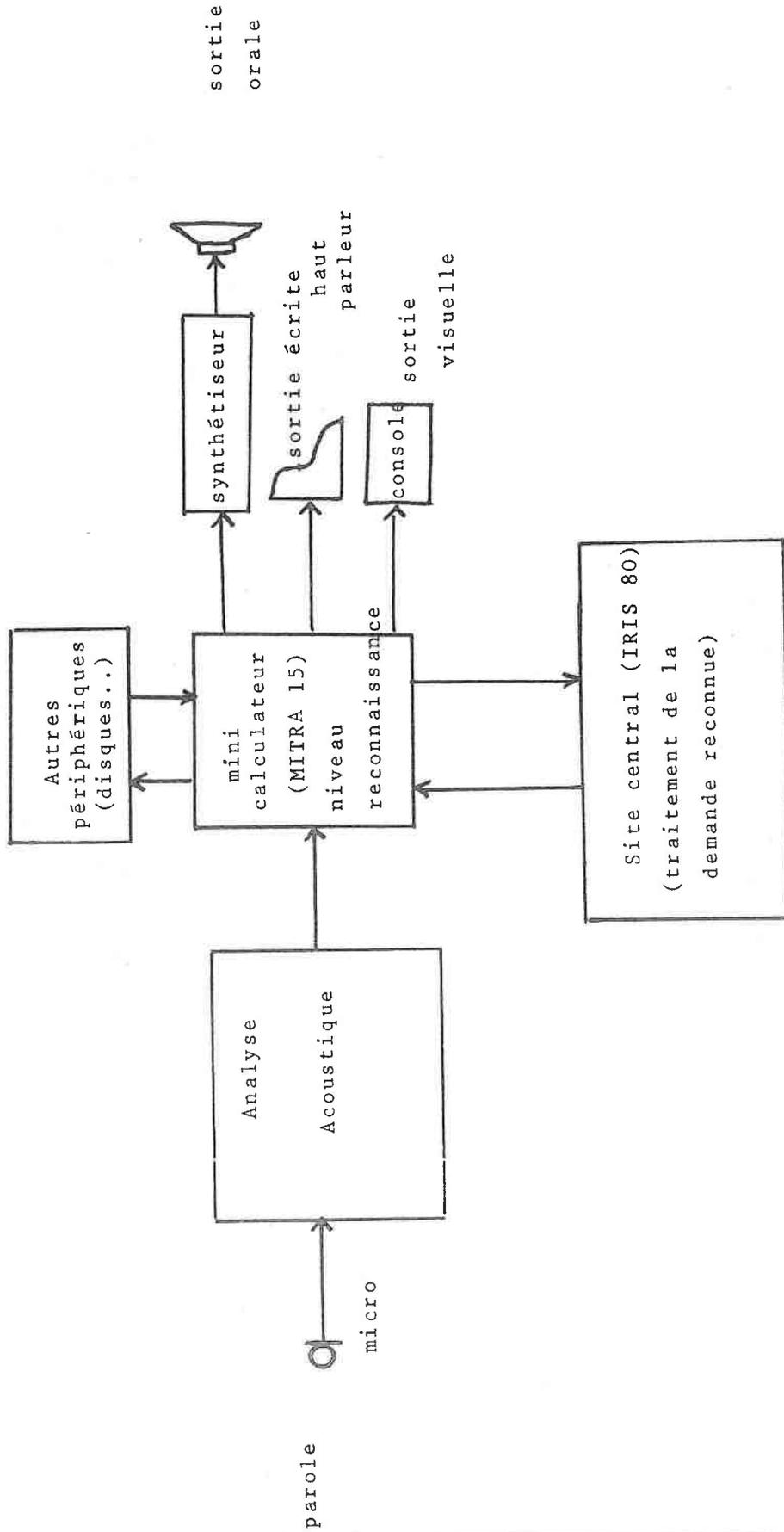
4-2- GENERALISATION DU SYSTEME :

4-2-1- Augmentation de la taille du vocabulaire accepté :

L'une des premières généralisations possibles du système présenté dans ce mémoire consiste à augmenter le nombre de symboles terminaux et ainsi à effectuer des tests sur des langages plus vastes. Ceci peut se faire soit par l'augmentation du nombre de réalisations des terminaux de type lexicque soit par l'accroissement de la grammaire de référence utilisée. Le premier cas n'agit que sur le déroulement du module de recherche lexicale et une telle généralisation peut être facilement implémentée sans modifier de façon sensible les performances du système, du moins dans le cas où le nombre de réalisations de ces terminaux reste inférieur à une certaine limite qui devra être déterminée par des tests successifs du module de recherche lexicale. Quant au second cas, il faut distinguer deux types d'accroissement possible du langage.

a) accroissement par substitution dans la grammaire.

Il correspond au cas où dans la grammaire du langage L_1 , on remplace un symbole, terminal ou non terminal, par l'axiome de la grammaire d'un langage L_2 de même type pour obtenir la définition d'un troisième langage L_3 . Une telle généralisation ne doit pas poser de problèmes importants, même si le nombre de phrases et de terminaux se trouve ainsi considérablement augmenté. La place mémoire nécessaire pour traiter le nouveau langage ainsi formé sera, certes, plus importante mais on peut espérer que l'efficacité du système



Implantation d'un système de compréhension de la parole

Figure C-4-2-

ne s'effondrera pas. En effet, à chaque instant, la reconnaissance se fait localement sur le graphe représentant la grammaire et, les retours en arrière étant peu nombreux et en général d'un seul pas, le temps de traitement nécessaire pour reconnaître une phrase du langage L3 doit rester proportionnel à la somme des temps obtenus pour L1 et L2.

b) accroissement par adjonction dans la grammaire.

Ceci revient à augmenter la complexité de la grammaire traitée en ajoutant de nouvelles possibilités au niveau de chaque noeud du graphe correspondant. Le nombre d'hypothèses émises à chaque instant par l'analyseur lexicographique se trouve ainsi considérablement augmenté. Dans ce cas, il semble qu'à partir d'un certain seuil les performances du système au niveau temps de calcul risquent de s'effondrer même si la longueur de la phrase varie peu. Seule une étude approfondie du système dans de tels cas pourrait nous permettre de déterminer ce seuil qui vraisemblablement doit être supérieur à plusieurs dizaines de possibilités pour chaque noeud.

De plus il serait intéressant de restreindre le nombre de comparaisons, effectuées lors de la recherche lexicale, pour augmenter la vitesse du processus de reconnaissance. Dans ce but, on peut envisager une partition du lexique en classant les mots en quelques grands types facilement différenciables :

- mots contenant une ou plusieurs occlusions
- mots contenant un ou plusieurs segments fricatifs.

Ainsi, au vu de la chaîne phonémique à traiter, la comparaison ne se ferait seulement qu'à l'intérieur de l'une de ces sous-classes et améliorerait sensiblement les performances du système.

4-2-2- Augmentation du nombre de locuteurs :

Pour l'instant notre système n'a été testé que pour un seul locuteur ou une classe unique de locuteurs ayant des voix suffisamment proches. Or, comme nous l'avons déjà noté, les applications grand-public utilisant par exemple le téléphone comme moyen de communication nécessitent une généralisation du système lui permettant de traiter des phrases énoncées par divers locuteurs, hommes, femmes et enfants. Nous n'avons pas du tout abordé ce problème, mais si l'on se réfère aux résultats fournis par d'autres systèmes, on constate une rapide détérioration des résultats lorsque le nombre de locuteurs augmente. Pour pallier cet inconvénient, il faudrait prévoir une procédure d'adaptation aux divers locuteurs qui pourrait être incluse dans une phase d'apprentissage :

4-3- PHASE D'APPRENTISSAGE :

Dans un système de reconnaissance automatique de la parole ou, plus généralement, de formes, plutôt que de fournir globalement les règles de classement des différentes formes à traiter, on procède à un ajustement progressif des paramètres, par présentation d'un grand nombre de formes. Cette phase, appelée apprentissage, est en général un processus séquentiel nécessitant des calculs importants. Le système se corrige peu à peu au cours du temps et à mesure que de nouvelles formes se présentent : il acquiert ainsi une sorte d'expérience.

Dans le cadre d'un système tel que celui que nous venons de présenter, on peut distinguer trois niveaux différents d'apprentissage possibles : phonémique, morphologique et syntaxique.

4-3-1- Au niveau phonémique :

Le premier niveau d'apprentissage correspond à la mise au point du système acoustique de reconnaissance phonémique. Comme nous l'avons vu (partie A ch. 2) les phonèmes sont reconnus grâce à divers paramètres (en particulier les fréquences des formants et du fondamental) et par comparaison avec des spectres de références représentant ces phonèmes.

On peut à ce niveau mettre en place un système permettant d'acquérir ces paramètres par évolutions successives. Les sons correspondant à chaque phonème émis séparément puis dans divers contextes peuvent permettre ainsi au système de s'adapter à chaque locuteur. Un tel apprentissage risque d'être très long et un travail intéressant pourrait consister à rechercher quelques phrases types du français parlé que la machine ferait répéter, pour procéder ensuite le plus rapidement possible aux ajustements nécessaires des paramètres permettant d'accepter un nouveau locuteur.

4-3-2- Au niveau morphologique :

Un second stade d'apprentissage doit se situer au niveau morphologique. En effet comme nous l'avons déjà vu (partie B chapitre 2) chaque mot du langage est stocké en mémoire dans un dictionnaire sous forme d'un graphe représentant la chaîne phonémique de référence. Il est, certes, souhaitable de prévoir au départ une représentation de référence théorique. Pourtant, au cours de l'étude de la recherche lexicale (partie B chapitre 5) nous avons vu que le système émettait des hypothèses d'insertion, de substitution ou d'élision de phonèmes. On peut en déduire qu'il serait intéressant pour le système de mémoriser les diverses hypothèses faites lors de la reconnaissance d'un mot ainsi que leur nombre d'apparitions. Ainsi, à partir d'une certaine fréquence d'apparition d'une altération, le système pourrait l'inclure dans le graphe de référence du terminal traité. De telles mises au point peuvent en effet permettre une adaptation plus rapide à diverses prononciations et apporter ainsi une meilleure efficacité au module de recherche lexicale et, par ce biais, à l'ensemble du système.

4-3-3- Au niveau syntaxique :

Un dernier stade d'apprentissage peut se situer au niveau syntaxique, et un des outils proposé consiste à appliquer des méthodes d'inférence grammaticale (67) . On peut définir

le problème ainsi : peut-on, à partir d'une suite de phrases d'un langage L, trouver une grammaire qui l'engendre ?

Mis à part le fait qu'un tel apprentissage nécessite au départ une reconnaissance sans support syntaxique, car on ne peut utiliser la grammaire que l'on infère, il est difficile de juger si une grammaire inférée est optimale vis à vis de l'échantillon de départ. Dans le cas de langage infini, l'échantillon quant à lui est toujours fini et ainsi, on se heurte continuellement au cours de l'inférence à un choix important: faut-il généraliser la grammaire ou, au contraire, la faire coller de très près aux phrases qui ont servi à la créer ?

Nous pensons, pour notre part, qu'il n'est pas possible d'utiliser des méthodes d'inférence grammaticale dans le cadre de l'apprentissage d'un système de compréhension du discours continu et qu'il faudra toujours fournir au départ la grammaire d'un langage noyau de référence. Par contre, comme nous l'avons vu lors de l'étude de l'Analyseur lexicographique (partie B chapitre 4) le système émet à certains moments des hypothèses d'insertion, de substitution ou d'élision de mots. On peut donc, comme dans le cas de l'apprentissage morphologique, mémoriser les hypothèses faites et leur nombre d'apparition. Et ainsi, à partir d'une certaine fréquence d'apparition de l'une d'elles, modifier la grammaire de manière à tenir compte immédiatement de cette possibilité dès le traitement syntaxique.

4-4- AUTRES POSSIBILITES :

4-4-1- Caractérisation sémantique d'un langage :

Parmi les autres possibilités de recherche, pouvant contribuer à la reconnaissance et à la compréhension de la parole, on peut citer le domaine très vaste de la caractérisation sémantique des langages. Jusqu'alors en informatique deux types d'outils formels ont été principalement développés

et le plus souvent dans le cas spécifique de langages de programmation : ce sont la caractérisation par attributs et celle par double grammaire.

Il semble que de telles méthodes sont encore lourdes pour une utilisation efficace dans le cadre de systèmes de reconnaissance de la parole. Malgré cela, un travail intéressant peut consister dans l'étude de leur application pour de tels systèmes et, s'il s'avérait que ces méthodes soient trop difficilement utilisables en reconnaissance, il serait nécessaire de rechercher de nouveaux outils permettant une caractérisation sémantique d'un langage orientée vers la reconnaissance. De tels travaux pourraient ouvrir ainsi la voie à une meilleure formalisation de l'ensemble du traitement nécessaire à la compréhension de la parole et à l'établissement d'un dialogue oral fructueux entre l'utilisateur et la machine.

4-4-2- Définition de langages oraux de communication Homme-machine

Il reste enfin à aborder le problème de la définition de langages oraux permettant une utilisation optimale des possibilités actuelles, ou prévisibles à court et moyen terme, de la reconnaissance vocale. Pour faciliter une bonne communication entre l'utilisateur et la machine, et fournir ainsi un service efficace, il est nécessaire que de tels langages puissent être utilisés en temps réel ou du moins que le temps de réponse des systèmes reste de l'ordre de quelques secondes.

La reconnaissance en temps réel d'une langue naturelle, comme le français, semble encore lointaine; c'est pourquoi un tel langage doit être conçu comme un langage artificiel ou comme un sous-ensemble d'une langue naturelle. On peut bien sûr le regretter mais, actuellement, seuls des langages conçus dans le contexte d'applications particulières et bien définies peuvent apporter une aide réelle à la communication Homme-machine.

Ces langages artificiels doivent posséder un certain nombre de spécifications tenant compte du caractère particulier de la communication orale ; ils doivent, pour des français, avoir une consonnance "française" destinée à limiter le plus

possible la variabilité au niveau de la prononciation. Ils ne doivent pas avoir de contraintes trop artificielles pour permettre à l'utilisateur de se soucier plus du sens de la phrase orale qu'il prononce que de la forme. De plus, les mots du langage (ou les symboles terminaux) doivent être facilement distinguables sur le plan phonétique afin de réduire au maximum le nombre d'erreurs introduites lors de la reconnaissance.

Quant aux langages, sous ensemble de langues naturelles, ils sont presque indispensables dans le cas d'applications "grand-public" utilisant par exemple le téléphone comme moyen de communication (centre de renseignement, standard téléphonique etc...).

Il s'agit, en fait, de trouver un compromis acceptable entre la facilité d'utilisation du langage et les performances possibles du système.

Jusqu'alors, les langages utilisés sont expérimentaux et c'est en particulier le cas pour celui que nous avons testé pour notre système (cf figure C-1-2-). Il serait très utile de mener une étude sur la définition de ces langages oraux et d'obtenir ainsi une liste de caractéristiques plus définies qui pourrait aider efficacement les utilisateurs lors de la mise en place d'applications orales.

La programmation est aussi souvent évoquée dans le cadre de la communication orale homme-machine. Il semble pourtant que l'entrée orale d'un programme complet ne présente pas d'intérêt. Dans le cadre de la programmation, seuls des "outils" de travail permettant de traiter certains problèmes particuliers peuvent être utiles : langages oraux de consultation de banque de données, de gestion de fichiers, de commande de machines ou d'ordinateurs, d'aide à la mise au point de programmes.

A vrai dire, la communication orale homme-machine n'est encore qu'au stade expérimental et la définition de normes concernant les langages oraux reste un domaine de recherche largement ouvert dont nous n'avons présenté ici que quelques aspects.

C O N C L U S I O N

L'orientation de nos recherches se place dans le cadre général de l'étude de la reconnaissance et de la compréhension du discours continu. Le travail présenté dans ce mémoire constitue une contribution à la mise en place de systèmes hiérarchisés de reconnaissance de la parole composés de divers niveaux : acoustique, phonétique, syntaxique, morphologique et sémantique.

Nos recherches ont porté plus spécialement sur l'étude des niveaux syntaxiques et morphologiques et nous ont amenés à la mise au point d'un système de compréhension guidé par un analyseur syntaxique descendant dont la structure a été présentée dans la seconde partie de ce mémoire. Outre l'utilisation de la syntaxe, les trois caractéristiques particulières de ce système sont :

- sa généralité : construit de manière paramétrable, il est en effet capable de traiter tous les langages à contexte libre et est ainsi facilement adaptable à diverses applications.

- son aptitude à prendre certaines libertés par rapport à la grammaire du langage noyau en paramètre : il autorise ainsi une plus grande souplesse pour l'utilisateur.

- son utilisation du dialogue : il permet en effet un véritable échange entre l'Homme et la machine et a la possibilité de corriger des erreurs ou des ambiguïtés de reconnaissance grâce au dialogue.

Dans la troisième partie nous avons présenté une application de ce système dans le cadre particulier de l'automatisation d'un standard téléphonique. Les résultats obtenus nous ont permis de valider l'approche utilisée et ont montré, de plus, qu'un tel système est capable de travailler en temps réel lors du traitement de langages spécialisés.

Nous avons enfin présenté quelques voies de recherches qui nous semblent intéressantes pour permettre une amélioration sensible de tels systèmes.

L'ensemble de ces travaux nous ont surtout permis de mieux cerner les différents problèmes qui se posent pour parvenir à la compréhension du discours continu. Si l'intérêt de la mise en place de tels systèmes est certain, il nous semble surtout nécessaire de prendre un certain recul par rapport aux réalisations particulières du type de celle présentée ici, afin d'étudier plus à fond les "outils" nécessaires à une meilleure formalisation du processus de reconnaissance et de compréhension du discours continu.

Le système que nous avons réalisé se place dans le cadre de l'étude des entrées vocales et plus généralement de l'amélioration de la communication Homme-machine. Nous espérons avoir contribué à cette amélioration et pouvoir continuer à travailler dans ce sens.

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

Les références citées apparaissent ici
dans leur ordre de première apparition dans le texte.

- (1) J.P. HATON "Contribution à l'analyse, la paramétrisation et la reconnaissance automatique de la parole"
Thèse d'Etat, Université de Nancy I,
Janvier 1974.
- (2) B. MALMBERG "La phonétique" Que sais-je ? n°637 P.U.F.
- (3) J.S. LIENARD "Analyse, Synthèse et reconnaissance automatique de la parole" thèse d'Etat,
Université de Paris VI, Avril 1972.
- (4) J.W. COOLEY, J.W. TUKEY
"An algorithm for the machine calculation of complex Fourier series" Math. Comput,
19, pp 297-301, April 1965.
- (5) E. PETERSON "Frequency detection and speech formants"
JASA, 23, pp 668-74, 1951.
- (6) P. VICENS "Aspects of speech recognition by computer"
Ph. D. Thesis-Stanford University, April 1969
Memo A.I. 85.
- (7) J. MAKHOUL "Spectral analysis of speech by linear prediction" Int. Conf. on speech com. and Proc.
Newton Mass. April 1972.
- (8) A.N. MUCCIARDI, E.E. GOSE
"A comparison of seven techniques for choosing subsets of pattern recognition properties"
IEEE Tr. Comp., 20, n°9, p 1023-1031, 1971.
- (9) J. DREYFUS-GRAF
"SONOGRAPH and sound Mechanics"
JASA, 22, pp731-39, 1950.
- (10) K.H. DAVIS et al.
"Automatique recognition of spoken digits"
JASA, 24, p 637, 1952.
- (11) H. DULEY, S. BALASHEK
"Automatic recognition of phonetic patterns in speech" JASA, 30, p 721-32, 1958.

- (12) H.F. OLSON, H. BELAR
"Phonetic typewriter" JASA, 28, pp 1071-81,
1956.
- (13) P. DENES "The design and operation of the mechanical
speech recognizer at University College,
London" J. Brit. Int. Radio Eng., 19,
p 219-20, 1959.
- (14) G.L. SHULZ "Investigation procedures for speech recogni-
tion" Proc. seminar on speech comp and Proc.
Cambridge Mass. sept. 1959.
- (15) J.H. KING, C.J. TUNIS
"Some experiments in spoken word recognition"
IBM Journal, 10, n°1, pp 65-79, 1966.
- (16) B. COLD "Word recognition computer program" Tech. Rept
n° 452, Lincoln Lab MIT, 1966.
- (17) D.R. REDDY "Segmentation of speech sounds"
JASA n°40, pp 307-12, 1966.
- (18) E.T. SEMENOVA
"Analysis of speech recognition criteria"
For. Tech. Dev. Wright Patterson AFB,
n°AD680-571, april 1968.
- (19) U.N. TRUNIN-DONSKOI
"Recognition of spoken word ensemble with the
aid of a digital computer" in "Works on
technique Cybernetics" Moscou URSS 1969.
- (20) D. DOURS-R. FACCA-G. MAURAND-G. PERENNOU
"L'apport de l'analyseur du projet ARIA sur
quelques exemples d'analyse phonétique"
6^e JEP, GALF, Toulouse, Mai 1975.
- (21) D. DOURS-R. FACCA
"Méthode de segmentation et d'analyse par
traitement direct du signal vocal"
Thèse Univ. P. Sabatier Toulouse 1974.

- (22) L. BUISSON, G. MERCIER
"Utilisation de l'information prosodique
en segmentation de la parole continue"
6^e SEP, GALF, Toulouse, Mai 1975.
- (23) D. TUFFELLI "Détermination de formants et de traits
par codage prédictif 6^e SEP, GALF,
Toulouse, Mai 1975.
- (24) R. ALTER "Utilisation of contextual constraints in
Automatique speech recognition" IEEE Trans
Audio and Electroacoustic vol. AU 16 n° 1,
p 6-11, 1968.
- (25) D.R. REDDY, A.E. ROBINSON
"Phonème to grapheme translation of English"
IEE Trans AU-16 n° 2, p240-46, 1968.
- (26) R.B. NEELY, G.M. WHITE
"On use of syntax in a low cost real time
speech recognition system "
Informating processing, 1974.
- (27) M.B. HERSCHER and R.B. COX
"An adaptive isolated word speech recognition
system"
Speech comm. and Proc. Boston 1972
- (28) D.R. REDDY et Al.
"The C.Mu speech recognition project"
Proc IEEE Trans. Syst. Sci. Cyb. Conf. 1971.
- (29) W.R. LESSER, R.D. FENNEL, L.D. ERMAN, D.R. REDDY
"Organisation of the HERSAY II speech unders-
tanding system" IEEE Symposium of speech
recognition CMU 1974.
- (30) D.R. REDDY, L.E. ERMAN, N.B. NEELY
"A model and a system for machine recognition
of speech" IEE Trans. on audio and electro-
acoustics vol AU 21 n°3 Juin 1973.
- (31) J.W. KLOWSTAT, L.F. MONSHEIM
"The CASPERS linguistic analysis system"
IEEE symposium of speech recognition CMU 1974.

- (32) P.L. MILLER "A locally organized parser for spoken input" Communication of the ACM
Vol 17 n° 11 1975.
- (33) J.K. BAKER "The Dragon system an overview" IEEE
symposium of speech recognition CMU 1974
- (34) J. BARNETT "A Vocal Data management system"
IEEE Trans. Audio and electroacoustics
Vol AU 21 n° 3 pp 185-188, Juin 1973.
- (35) H.B. RITEA "Speech input to a data management system"
Speech communication seminar, Vol 3
pp 291-298, Stockholm Aug 1-3, 1974.
- (36) C.C. TAPPERT, N.R. DIXON, A.S. RABINOWITZ
"Application of sequential decoding for
converting Phonetic to graphic representation
in Automatic Recognition of Continuous Speech
(ARCS)" IEEE Trans an Audio and Electroacous-
tics Vol AU 21 n°3 Juin 73
- (37) G.Y. VYSOTSKY et al
"an experience in oral control of a computer"
Eng Cybern n°2, p 320-27, 1970.
- (38) M. CARRRIER, R. CARRE, J.Y. GRESSER
"Compte rendu d'une mission d'étude en URSS :
analyse, synthèse, reconnaissance automatique
et perception de la parole" dec 1971, CEI/7,
TRAM/5.
- (39) J.P. TUBACH "Reconnaissance automatique de la parole"
Thèse d'Etat, Université de Grenoble 1970.
- (40) J.P. HATON "A practical application of a real time isolated
word recognition system using syntactic cons-
traints" IEEE Trans ASSP, 22, n°6, p 416-419,
1974.
- (41) J.P. HATON, M.C.HATON, M.LAMOTTE
"Syrène, un système interactif pour la réedu-
cation vocale des non entendants" 6^e JEP,
GALF, TOULOUSE, Mai 1975.

- (42) Laboratoires BELL
 "Experimental Voice controlled device gives" "command" "performances" JASA, 50, n° 6, p 1450, 1971.
- (43) J.P. HATON "Acoustic segmentation and real-time recognition of speech" 8^e congrès international d'Acoustique Londres Juillet 1974.
- (44) R. BELLMAN "Dynamic programming" Princeton Univ. Pres. 1957.
- (45) J.P. HATON "The use of dynamic algorithms in automatic speech recognition" 7^e Int. Cong. of Cybernetique Namur 7-11 Spet. 1973.
- (46) J.P. HATON "Rôles des contraintes linguistiques en reconnaissance de la parole" Journées d'Etude sur la parole GALF Lannion Juillet 1972.
- (47) I.EL. MALLAWANY
 "Contribution aux recherches sur la communication parlée" Thèse de docteur ingénieur, Université de Grenoble, septembre 1975.
- (48) C. PAIR "Compilation" école d'été d'informatique AFCET, Neufchatel, Juillet 1972.
- (49) C. PAIR "Structures de données et algorithme fondamentaux" Cours ENSMIM, 1974, NANCY.
- (50) J.M. PIERREL, J.P. HATON
 "Une approche syntaxique de reconnaissance de phase dans un contexte donné" 6^e JEP, GALF, Toulouse, Mai 1975.
- (51) QUINTON P., VIVES, GRESSER J.Y.
 "Dialogue avec un robot" 6^e JEP, GALF, Toulouse, Mai 1975.
- (52) C. PAIR "Analyse syntaxique" Ecole d'Eté EDF-CEA-IRIA 1973
- (53) S.A. GREIBACH
 "A new normal form theorem for context free phrase structure grammars" Journal ACM 12 : 1 pp 42-52 1965.

- (54) J.C. DERNIAME, C. PAIR
"Problème de cheminement dans les graphes"
Dunod, 1973
- (55) WARSHALL "A theorem on booleen matrices", J of ACM
Janvier 1962.
- (56) A.J. SZANSER "Elastic matching in automatic pattern recognition" Proc. int. Conf. on Mach. Perc. of patterns and pictures, Teddington, G.B.,
April 1972.
- (57) R. VIVES "Elaboration d'un indice de ressemblance entre deux chaines de longueur finie quelconque"
Note technique CEI/CSI/27 CNET Oct. 1972.
- (58) R.M. FANO "A heuristic discussion of probabilistic decoding" IEEE Tr. on Inf. Theory, 9, pp 64-73
1964.
- (59) A.S. RABINOWITZ
"On the value of the Fano algorithm in establishing the graphemic form of machine derived phonetic strings" Th D. Thesis,
Univ. of North Carolina, Raleigh, March 1972.
- (60) G.BATTANI et H. MELONI
"Mise en oeuvre de contraintes phonologiques syntaxiques et sémantiques dans un système de compréhension automatique de la parole"
Thèse, Université d'Aix-Marseille, Juin 1975.
- (61) C. GUEGEN et al.
"Un synthétiseur à structure programmable"
6^e JEP, GALF, Toulouse, Mai 1975.
- (62) J. GREVIN, F. BERTIN, J. SAP
"Synthétiseur de parole à formants à circuits numériques" 6^e JEP, GALF, Toulouse, Mai 1975.
- (63) X. RODET et C. SANTAMARINA
"Syrus : synthèse, sur un mini-ordinateur, du signal de parole dans sa représentation amplitude temps" 6^e JEP, GALF, Toulouse,
Mai 1975.

- (64) J.L. COURBON "Simulation du conduit vocal en technologies analogiques" 6^e JEP, Toulouse, Mai 1975.
- (65) J.P. HATON et J.M. PIERREL
"Essai d'automatisation d'un standard téléphonique à l'aide d'un système de reconnaissance de la parole continue"
Colloque n° 1 l'Acoustique dans les télécommunications -F.A.S.E. 75-Paris Octobre 1975.
- (66) C.C. TAPPERT and N.R. DIXON
"A procedure for adaptive control of the interaction between acoustic classification and linguistic decoding in automatic recognition of continuous speech" Artificial intelligence n° 5 1974 pp 95-113.
- (67) L. MICLET "Inférence automatique de grammaires formelles" 6^e journées d'étude sur la parole, GALF Toulouse, Mai 1975.
- (68) V.A. AHO et J.D. ULLMAN
"The Theory of parsing, Translation and compiling". Vol 1 Prington Hall 1972
Vol 2 Prington Hall 1973.
- (69) D.E. KNUTH "Top down syntax analysis" Acta informatica, 1, pp 79-110, 1971

NOM DE L'ETUDIANT : PIERREL J.M.

NATURE DE LA THESE : DOCTORAT DE SPECIALITE EN MATHEMATIQUES

VU, APPROUVE

& PERMIS D'IMPRIMER

NANCY, le 13 NOV. 1975

LE PRESIDENT DE L'UNIVERSITE DE NANCY I

Helloy
UNIVERSITÉ DE NANCY
Le Président

J.R. HELLOY