



HAL
open science

Transcription orthographique-phonétique : un système interpréteur

Monique Lety

► **To cite this version:**

Monique Lety. Transcription orthographique-phonétique : un système interpréteur. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1980. Français. NNT : . tel-00293089

HAL Id: tel-00293089

<https://theses.hal.science/tel-00293089>

Submitted on 3 Jul 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

présentée à

l' Université Scientifique et Médicale de Grenoble

pour obtenir le grade de

DOCTEUR INGENIEUR

(Informatique)

par

Monique LETY



TRANSCRIPTION ORTHOGRAPHIQUE – PHONETIQUE :

UN SYSTEME INTERPRETEUR



Thèse soutenue le 27 novembre 1980 devant la Commission d'Examen :

Monsieur	G. VEILLON	: Rapporteur
Monsieur	J. COURTIN	: Président
Messieurs	L.J. BOË	} Examineurs
	J. GENIN	
	V. LUCCI	

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Monsieur Gabriel CAU : Président

Monsieur Joseph KLEIN : Vice-Président

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

PROFESSEURS TITULAIRES

MM.	AMBLARD Pierre	Clinique de dermatologie
	ARNAUD Paul	Chimie
	ARVIEU Robert	I.S.N.
	AUBERT Guy	Physique
	AYANT Yves	Physique approfondie
Mme	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique expérimentale
	BARBIER Reynold	Géologie appliquée
	BARJON Robert	Physique nucléaire
	BARNOUD Fernand	Biosynthèse de la cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique chirurgicale A
	BEAUDOING André	Clinique de pédiatrie et puériculture
	BELORIZKY Elie	Physique
	BARNARD Alain	Mathématiques pures
Mme	BERTRANDIAS Françoise	Mathématiques pures
MM.	BERTRANDIAS Jean-Paul	Mathématiques pures
	BEZES Henri	Clinique chirurgicale et traumatologie
	BLAMBERT Maurice	Mathématiques pures
	BOLLIET Louis	Informatique (I.U.T. B)
	BONNET Jean-Louis	Clinique ophtalmologie
	BONNET-EYMARD Joseph	Clinique hépato-gastro-entérologie
Mme	BONNIER Marie-Jeanne	Chimie générale
MM.	BOUCHERLE André	Chimie et toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques appliquées
	BOUTET DE MONVEL Louis	Mathématiques pures
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique rhumatologique et hydrologique
	CALAS François	Anatomie
	CARLIER Georges	Biologie végétale
	CARRAZ Gilbert	Biologie animale et pharmacodynamie

MM.	CAU Gabriel	Médecine légale et toxicologie
	CAUQUIS Georges	Chimie organique
	CHABAUTY Claude	Mathématiques pures
	CHARACHON Robert	Clinique ot-rhino-laryngologique
	CHATEAU Robert	Clinique de neurologie
	CHIBON Pierre	Biologie animale
	COEUR André	Pharmacie chimique et chimie analytique
	COUDERC Pierre	Anatomie pathologique
	DEBELMAS Jacques	Géologie générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumophtisiologie
	DEPORTES Charles	Chimie minérale
	DESRE Pierre	Métallurgie
	DODU Jacques	Mécanique appliquée (I.U.T. I)
	DOLIQUE Jean-Michel	Physique des plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	FONTAINE Jean-Marc	Mathématiques pures
	GAGNAIRE Didier	Chimie physique
	GALVANI Octave	Mathématiques pures
	GASTINEL Noël	Analyse numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique générale
	KLEIN Joseph	Mathématiques pures
	KOSZUL Jean-Louis	Mathématiques pures
	KRAVTCHENKO Julien	Mécanique
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie végétale
Mme	LAJZEROWICZ Janine	Physique
MM.	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie générale
	LATURAZE Jean	Biochimie pharmaceutique
	LAURENT Pierre	Mathématiques appliquées
	LEDRU Jean	Clinique médicale B
	LE ROY Philippe	Mécanique (I.U.T. I)

MM.	LLIBOUTRY Louis	Géophysique
	LOISEAUX Jean-Marie	Sciences nucléaires
	LONGEQUEUE Jean-Pierre	Physique nucléaire
	LOUP Jean	Géographie
Mlle	LUTZ Elisabeth	Mathématiques pures
MM.	MALINAS Yves	Clinique obstétricale
	MARTIN-NOEL Pierre	Clinique cardiologique
	MAYNARD Roger	Physique du solide
	MAZARE Yves	Clinique Médicale A
	MICHEL Robert	Minéralogie et pétrographie
	MICOUD Max	Clinique maladies infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie nucléaire
	NEGRE Robert	Mécanique
	NOZIERES Philippe	Spectrométrie physique
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques pures
	PEBAY-PEYROULA Jean-Claude	Physique
	PERRET Jean	Séméiologie médicale (neurologie)
	RASSAT André	Chimie systématique
	RENARD Michel	Thermodynamique
	REVOL Michel	Urologie
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-Chirurgie
	SARRAZIN Roger	Clinique chirurgicale B
	SEIGNEURIN Raymond	Microbiologie et hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction mécanique (I.U.T. I)
	SOUTIF Michel	Physique générale
	TANCHE Maurice	Physiologie
	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique nucléaire
Mme	VERAIN Alice	Pharmacie galénique
MM.	VERAIN André	Physique biophysique
	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale

PROFESSEURS ASSOCIES

MM. CRABBE Pierre
SUNIER Jules

CERMO
Physique

PROFESSEURS SANS CHAIRE

Mlle	AGNIUS-DELORS Claudine	Physique pharmaceutique
	ALARY Josette	Chimie analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	ARMAND Gilbert	Géographie
	BENZAKEN Claude	Mathématiques appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique (I.U.T. I)
	BUISSON René	Physique (I.U.T. I)
	BUTEL Jean	Orthopédie
	COHEN-ADDAD Jean-Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie médicale
	CONTE René	Physique (I.U.T. I)
	DELOBEL Claude	M.I.A.G.
	DEPASSEL Roger	Mécanique des fluides
	GAUTRON René	Chimie
	GIDON Paul	Géologie et minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biochimie médicale
	HACQUES Gérard	Calcul numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et médecine préventive
	IDELMAN Simon	Physiologie animale
	JOLY Jean-René	Mathématiques pures
	JULLIEN Pierre	Mathématiques appliquées
Mme	KAHANE Josette	Physique
MM.	KRAKOWIACK Sacha	Mathématiques appliquées
	KUHN Gérard	Physique (I.U.T. I)
	LUU DUC Cuong	Chimie organique - pharmacie
	MICHOULIER Jean	Physique (I.U.T. I)
Mme	MINIER Colette	Physique (I.U.T. I)

MM.	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et minéralogie
	PFISTER Jean-Claude	Physique du solide
Mlle	PIERY Yvette	Physiologie animale
MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REYMOND Jean-Charles	Chirurgie générale
	RICHARD Lucien	Biologie végétale
Mme	RINAUDO Marguerite	Chimie macromoléculaire
MM.	SARROT-REYNAULD Jean	Géologie
	SIROT Louis	Chirurgie générale
Mme	SOUTIF Jeanne	Physique générale
MM.	STIEGLITZ Paul	Anesthésiologie
	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	ARMAND Yves	Chimie (I.U.T. I)
	BACHELOT Yvan	Endocrinologie
	BARGE Michel	Neuro-chirurgie
	BEGUIN Claude	Chimie organique
Mme	BERIEL Hélène	Pharmacodynamie
MM.	BOST Michel	Pédiatrie
	BOUCHARLAT Jacques	Psychiatrie adultes
Mme	BOUCHE Liane	Mathématiques (CUS)
MM.	BRODEAU François	Mathématiques (I.U.T. B) (Personne étrangère habilitée à être directeur de thèse)
	BERNARD Pierre	Gynécologie
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie papetière
	CHIAVERINA Jean	Biologie appliquée (EFP)
	COLIN DE VERDIERE Yves	Mathématiques pures
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	CORDONNER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie

MM.	CYROT Michel	Physique du solide
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie végétale
	DUSSAUD René	Mathématiques (CUS)
Mme	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FAURE Gilbert	Urologie
	GAUTIER Robert	Chirurgie générale
	GIDON Maurice	Géologie
	GROS Yves	Physique (I.U.T. I)
	GUIGNIER Michel	Thérapeutique
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	JALBERT Pierre	Histologie
	JUNIEN-LAVILLAVROY Claude	O.R.L.
	KOLODIE Lucien	Hématologie
	LE NOC Pierre	Bactériologie-virologie
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et médecine préventive
	MALLION Jean-Michel	Médecine du travail
	MARECHAL Jean	Mécanique (I.U.T. I)
	MARTIN-BOUYER Michel	Chimie (CUS)
	MASSOT Christian	Médecine interne
	NEMOZ Alain	Thermodynamique
	NOUGARET Marcel	Automatique (I.U.T. I)
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (I.U.T. B) (Personnalité étrangère habilitée à être directeur de thèse)
	PEFFEN René	Métallurgie (I.U.T. I)
	PERRIER Guy	Géophysique-glaciologie
	PHELIP Xavier	Rhumatologie
	RACHALL Michel	Médecine interne
	RACINET Claude	Gynécologie et obstétrique
	RAMBAUD Pierre	Pédiatrie
	RAPHAEL Bernard	Stomatologie
Mme	RENAUDET Jacqueline	Bactériologie (pharmacie)
MM.	ROBERT Jean-Bernard	Chimie-physique
	ROMIER Guy	Mathématiques (I.U.T. B) (Personnalité étrangère habilitée à être directeur de thèse)
	SAKAROVITCH Michel	Mathématiques appliquées

MM. SCHAERER René	Cancérologie
Mme SEIGLE-MURANDI Françoise	Crytogamie
MM. STOEBNER Pierre	Anatomie pathologie
STUTZ Pierre	Mécanique
VROUSOS Constantin	Radiologie

MAITRES DE CONFERENCES ASSOCIES

MM. DEVINE Roderick	Spectro Physique
KANEKO Akira	Mathématiques pures
JOHNSON Thomas	Mathématiques appliquées
RAY Tuhina	Physique

MAITRE DE CONFERENCES DELEGUE

M. ROCHAT Jacques	Hygiène et hydrologie (pharmacie)
-------------------	-----------------------------------

Fait à Saint Martin d'Hères, novembre 1977

TABLE DES MATIERES

	<u>Pages</u>
INTRODUCTION	
Chapitre I	Situation actuelle 6
	1 ^{re} Partie : ETUDE 13

Chapitre II	1- La grammaire 14
	1-1 Conception et formalisation des règles 14
	1-2 Symboles pouvant figurer dans le texte d'une règle : 15
	1-3 Analyse syntaxique d'une règle 23
	2- Fonctionnement de la grammaire 31
	2-1 Validation d'une règle de transcription 31
	2-2 Partitionnement et classement des règles 32
	2-3 Méthode de transcription d'un mot isolé 38
	3- Correction des erreurs 40
	3-1 Signalisation d'une erreur 40
	3-2 Localisation, dans la grammaire, de la provenance d'une erreur 40
	3-3 Résultats après correction 43
	4- Modifications et renseignements 44
Chapitre III	1- Transcription d'un texte 47
	1-1 Création du fichier contenant le texte à trans- crire 47
	1-2 Choix de la grammaire de transcription 49
	1-3 Transcription 50
	1-3-1 Méthode de transcription d'un texte 50
	1-3-2 Définition de mot et de nombre 51
	1-3-3 Transcription d'un mot 52
	1-3-4 Transcription d'un nombre 52
	1-3-5 Liaison-élision-dénasalisation 55
	1-3-6 Erreurs 58
	1-3-7 Présentation des résultats 58

	2- Etude statistique des règles de transcription	62
	2-1 Présentation	62
	2-2 Utilisation	64
	2-3 Application particulière	68
	Conclusion de l'étude : diversification des résultats	74
	1- Syllabation	74
	2- Transcription syllabique.....	77
	3- Transcription "libre".....	78
	 2 ^e Partie : REALISATION INFORMATIQUE	 80
Chapitre IV	----- 1- Dimensions de la grammaire	82
	2- Passage d'une règle formalisée à une règle codée.....	88
	3- Partitionnement et classement des règles	94
	4- Utilisation de la grammaire après une modification quelconque	98
	5- Passage d'une règle codée à la règle formalisée correspondante	98
	6- Mise en mémoire d'une grammaire.....	99
	7- Transcription d'un mot	99
	8- Renseignements et modifications.....	102
Chapitre V	1- Examen du texte à transcrire.....	103
	1-1 Présentation du texte	103
	1-2 Traitement du texte	104
	1-3 Particularités relatives à la nature de la transcription.....	111
	2- Statistiques élémentaires	111
	 CONCLUSION	 113
	ANNEXE : -conversion nombre-graphèmes.....	117
	-mode d'emploi du programme MOD7F	121
	GLOSSAIRE : termes linguistiques	129
	BIBLIOGRAPHIE	130

ERRATA

- p. 17 ligne 23 (cf. tableau p. 12).
- p. 19 3 lignes avant la fin : 35 phonèmes
- p. 22 5e ligne (en particulier ex 2 - 9)
- p. 23 ligne 13 : des 16 erreurs
- p. 33 §2.2.2. 3e ligne : on se bornerait
- p. 34 3° lère ligne : d'une règle ou LG
- p. 40 3 - 7e ligne : au lieu de "le tableau" lire la grammaire
- p. 40 3-2) 3e ligne sur la console
- p. 43 2e ligne la règle suivante :
- p. 44 4 - 3e ligne : il était souhaitable
- p. 49 b) 4e ligne (U) : savoir ⊂ ou 2
dernière ligne : mette
- p. 50 1.3.1. ligne 12 : 1967
- p. 59 d) 9e ligne : dans des fichiers
- p. 64 2-2) 5e ligne (cf. p. 45-46, chap. II)
- p. 68 6e ligne : d'un texte (cf. p. 65, 66, 67).
- p. 74 a) 4e ligne p. 20, ch. II.
- p. 76 2e ligne : § 1.3.7., chap. III), 5e ligne : § 1.3a, ch. V).
- p. 77 2°) lère ligne (cf. § 1.3.7a)
- p. 86 dernière ligne du 1er § : (au minimum)
- p. 92 4e tableau - 4e ligne : $15 \leq I \leq 38$
- p. 94 Tableau, 2e ligne : $6 \leq I \leq 5 + \dots$
- p. 99 3e ligne : la forme codée des règles qui peut être utile
dans la mise au point des programmes : ...
- p. 101 lère ligne : dont le plus long est ...
- p. 106 3e §, ligne 5 : que l'on désire ...
- p. 108 a rajouter à la suite du premier § :

1	K	50
1er mot	1	2° mot

- p. 108 Avant-dernier § : 2e ligne : va vérifier dans le texte
- p. 110 Programme : 4e ligne : ICARAC (1) ...
- 19e ligne : si ((CARLU=ponctuation et (N=0)) alors .
- 26e ligne : ORTPHO(ICARAC, 2)
- 34e ligne : ORTPHO(ICARAC,K)
- p. 113 a) 2e ligne : auraient pu être ...
- p. 114 c) 4e ligne : de cas possibleses ...

REMERCIEMENTS

-à M.VEILLON,pour m'avoir permis de disposer d'une allocation D.G.R.S.T.,ce qui m'a sans aucun doute aidée à m'en sortir,pour avoir pris de son temps déjà bien rempli afin de répondre à mes questions,pour m'avoir soutenue au cours de diverses démarches administratives...

-à L.J.BOË,pour son enthousiasme,son dynamisme,son humanité;pour la confiance qu'il m'a faite;pour la liberté qu'il laisse à ceux qui travaillent à l'INSTITUT de PHONETIQUE;pour avoir profité d'une ambiance que lui et les autres membres de son équipe ont instaurée à l'INSTITUT...

-à M.COURTIN et M.GENIN pour accepter de participer à ce jury...

-à M. LUCCI,pour s'être intéressé à mon travail;pour m'avoir donné les renseignements de linguistique qui me faisaient défaut et m'avoir ainsi évité de perdre du temps à les rechercher...

-à tous les professeurs dont j'ai suivi les cours à l'E.N.S.I.M.A.G.,et en particulier à M. SAKAROVITCH,pour leur compréhension et leur aide...

-à D.VUILLET,pour être toujours présente avec le même sourire,la même disponibilité le même savoir-faire...

-à Eric BOGNAR,Christian LEVEAU et d'autres encore,car l'amitié ça compte...

-à M.GUILLERMOU,Mme VERNAC,M.MONTEL pour m'avoir réconfortée quand il le fallait durant les trois années passées à l'E.N.S.I.M.A.G....

-à tous les membres de l'INSTITUT de PHONETIQUE,véritable équipe qui sait accueillir...

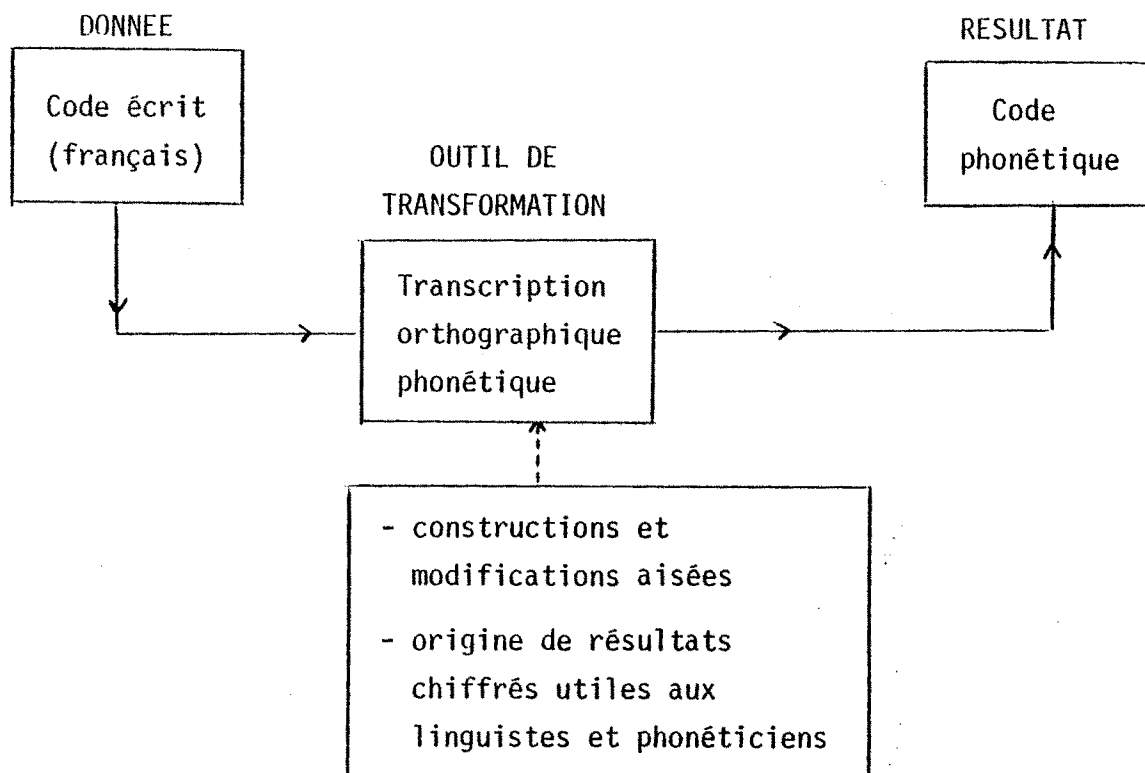
-à Mme HEIM et Mme BILDE,pour avoir rapidement dactylographié un manuscrit sans doute difficile à déchiffrer...

à ceux que j'ai nommés et à tous les autres je dirai que ces remerciements ne sont pas des remerciements de circonstance,mais je les fais pour la simple raison que si personne ne m'avait aidée,ça n'aurait jamais marché...

INTRODUCTION

Le sujet du travail présenté dans ce mémoire concerne la transcription orthographique - phonétique automatique. Cette appellation peut nous être tout à fait étrangère, par contre, le fait en lui-même nous est très familier : en effet, nous faisons inconsciemment cette transcription lorsque nous lisons un texte à voix haute. Transposé dans le domaine de l'informatique, ceci revient à réaliser la synthèse automatique de la parole à partir d'un texte écrit. Une transcription orthographique - phonétique est l'étape intermédiaire entre la reconnaissance des caractères dans le texte écrit et l'émission de signaux sonores. D'autre part, le lecteur d'un texte a aussi (entre autres) la possibilité de modifier sa diction, de corriger ses erreurs, d'apprendre à prononcer de nouveaux mots : il peut, en quelque sorte, modifier à volonté cette étape intermédiaire. Mais certaines personnes, font ce travail de transcription consciemment, afin d'étudier la nature et le fonctionnement de l'orthographe. Ceci est souvent long et fastidieux, et aboutit généralement à des comptages de graphèmes ou de phonèmes.

Ne serait-il pas possible de réunir dans un même programme de transcription orthographique - phonétique, les éléments permettant de réaliser, de façon automatique, ce que les uns font inconsciemment et d'autres consciemment ? Schématiquement, on peut situer ainsi cette étape de transcription :



Le travail à réaliser est donc constitué de la conception et de la mise au point d'un tel outil.

Réaliser sur ordinateur une toute petite partie de ce qu'un cerveau humain nous permet d'effectuer d'une façon inconsciente peut paraître à la fois ambitieux et mystérieux. Mais chacun de nous a appris à lire : on nous a, en quelque sorte, mis en mémoire une grammaire de transcription orthographique - phonétique, que l'habitude rend par la suite inutile. Mais, apprendre à lire pose tout de même certaines difficultés : il faut comprendre ce qui est lu ! L'analogie entre ce programme de transcription orthographique - phonétique, et le fait de lire un texte à voix haute se bornera donc en l'identité de ces résultats phonétiques qui sont dans le premier cas des résultats écrits, et dans le second des résultats sonores.

Malgré cette restriction, de nombreuses difficultés demeurent dans la réalisation d'un programme de transcription orthographique - phonétique. Quelles sont-elles, d'où proviennent-elles ? La donnée principale du problème, c'est-à-dire un texte écrit, en contient la cause majeure : l'orthographe. En ce qui concerne le travail de transcrip-

tion, le jugement de Voltaire prévaut : "L'écriture est la peinture de la voix; plus elle est ressemblante, meilleure elle est" (cf. Thimonnier p.294).

Mais ceci n'est pas le cas actuellement, pour la langue française. Aussi naissent de nombreux projets de réforme de l'orthographe auxquels répondent les défenseurs de cette dernière. Voici, dans l'un et l'autre cas, des arguments qui laissent entrevoir une possibilité d'automatisation d'une transcription, d'autres qui en relèvent la difficulté, et enfin, un point d'accord entre partisans et détracteurs de l'orthographe française actuelle.

a) - Possibilité d'automatisation

V. GAK
p. 31

"... Observons que le changement de valeur des lettres, isolées ou groupées, est soumis à des règles assez strictes (...) La possession des principaux de ces groupes et des principales variations de signification déterminées par leur entourage permet de déchiffrer 95% des mots ..."

R. THIMONNIER
p. 306

" ... Enfin si chacun sait qu'en français moderne, tout ce qui s'écrit ne se prononce pas, on oublie de signaler que tout ce qui se prononce doit s'écrire ..."

C.B. BENVENISTE &
A. CHERVEL

" ... On doit bien admettre cependant que l'écriture française est fondée, qu'elles qu'en soient l'insuffisance et les variations au cours des âges, sur un ensemble de correspondance entre des sons et des lettres. Entre le système phonologique actuel et le matériau graphique se sont établis un certain nombre de relations plus ou moins régulières qui constituent le code phonographique ..."

b) - Difficultés d'une transcription automatique

- V. GAK
p. 24 - 25
- "... L'idéal de l'écriture alphabétique serait qu'à chaque son corresponde une lettre distincte et que, réciproquement, chaque lettre ne possède qu'une seule valeur phonique. En fait une telle concordance s'observe rarement dans les systèmes graphiques existants. Mais on peut dire (en forçant un peu) qu'en français l'inobservation de ce principe est de règle ..."
- R. THIMONNIER
p. 295
- "... C'est que par nature, la langue parlée évolue sans cesse tandis que la langue écrite tend à se fixer. Il est donc normal que, dans les langues à évolution rapide (cas de l'anglais et du français), l'écriture et la parole apparaissent comme deux réalités plus ou moins indépendantes ..."
- C.B. BENVENISTE &
A. CHERVEL
p. 13
- "... En particulier, il est établi que toutes les écritures de type phonétique accordent une partie si minime fût-elle, à l'idéographie ..."
- V. LUCCI &
Y. NAZE
p. 33
- "... L'écriture est toujours plus ou moins indépendante de la langue l'orthographe étant en partie conventionnelle ..."

c)- Point d'accord

- V. GAK
p. 70
- "... L'écriture du français restant à base phonétique ..."
- R. THIMONNIER
p. 298
- "... Chaque fois qu'une langue a une forme parlée et une forme écrite, on peut prétendre en effet que son écriture est directement ou indirectement phonétique ..."

C.B. BENVENISTE &
A. CHERVEL
p. 40

"... Elle implique donc que l'écriture alphabétique est faite pour transcrire des phonèmes ..."

V. LUCCI &
Y. NAZE
p. 74

"... On a déjà vu que notre écriture est fondamentalement phonologique ..."

N. CATACH
J. GOLFAND
R. DENUX
p. 74 - 75

"... Notre graphie est (...) d'ordre fondamentalement phonétique, ou plutôt phonologique ..."

CHAPITRE I

SITUATION ACTUELLE

La réalisation d'un programme de transcription orthographique - phonétique n'est pas une nouveauté en 1980. En présentant six méthodes et programmes déjà existants, il s'agit ici de dégager les points originaux de ce travail.

1.

"A procedure which uses a computer for transcribing french text into phonetic symbols" (PHONTRS)

Bruce Pratt and Georgette Silva

"Système de transcription orthographique - phonétique d'un texte français" (PHONTR)

P. Goyer, D. Degryse, B. Guérin

a) - Méthode : traitement du texte en 5 étapes

- syllabation du texte

- transcription en 4 phases successives : transcription des groupes valides (i. e. ne contenant pas de caractères phonétiques), de :

- 4 caractères ou plus

- 3 caractères

- 2 caractères

- 1 caractère,

au moyen de comparaisons aux éléments de listes, qui sont de même longueur que les groupes à transcrire.

b) - Buts : - synthèse de la parole à partir d'un texte écrit

- économie en temps et en place mémoire

- souplesse
- exactitude

- c) - Matériel : - PHONTRS : CDC 3200 (16 koctets de mémoire)
- PHONTR : PDP 11/20

2.

"Programme de transcription orthographique - phonétique en langue française"

H. Fervers, J. Le Roux, L. Miclet

- a) - Méthode : - Traitement du texte à transcrire en une étape
- On parcourt le texte à transcrire au moyen d'une fenêtre pouvant englober dix caractères. Le premier caractère non transcrit est toujours le quatrième caractère de cette fenêtre. On recherche le ou les phonèmes convenant à ce caractère, ou à un groupe de caractères successifs (à partir du 4ème) dans une grammaire contextuelle et arborescente. Puis la fenêtre avance dans le texte du nombre de caractères adéquats.
- b) - Buts : - synthèse de la parole à partir d'un texte écrit
- synthèse en temps réel
- matériel peu encombrant
- c) - Matériel : microprocesseur TMS 9900.

3.

"Pour une meilleure formalisation de la conversion automatique graphème - phonème"

N. Catach, V. Meissonnier

- a) - Méthode : Traitement du texte en 4 phases
- codification des fins de mot muettes

- traitement des liaisons (références à la syntaxe)
 - syllabisation des mots
 - phonétisation
- b) - Buts :
- synthèse de la parole à partir d'un texte écrit
 - en vu d'application pour les :
 - linguistes
 - acousticiens
 - psychologues
 - pédagogues
 - sociologues ...
 - compréhension de l'utilisateur
- c) - Matériel : Inconnu au vu des documents consultés.

4.

"Conception et réalisation sur ordinateur d'un programme de transcription graphème - phonétique du français"

M. Divay & M. Guyomard

- a) - Méthode :
- Traitement du texte à transcrire en 8 phases
 - traitement de la ponctuation et des blancs
 - conversion nombre - graphèmes
 - liaisons, enchaînements (1ère partie)
 - transcription mot à mot
 - suppression des géminées
 - liaisons, enchaînement (2ème partie)
 - élisions
 - passage à un mode externe.
- b) - Buts :
- transcription des phrases du français en vue de la synthèse
 - modifications et adaptation aisées des règles de transcription
 - outil de recherche en ce qui concerne les règles
- c) - Matériel : IRIS 80

5.

"Système de génération des phrases phonétiques"

G. Tep

- a) - Méthode : Traitement du texte en 7 phases
- découpage de la phrase graphique en mots
 - préconversion (concerne la ponctuation)
 - conversion à l'aide d'un lexique de 2000 mots, contenant des informations sur :
 - la catégorie syntaxique
 - le nombre, le genre
 - la personne (pour les verbes)
 - la possibilité de liaisons antérieures et postérieures
 - un critère sémantique
 - la possibilité d'une élision
 - les racines graphique et phonétique
 - la désinence
 - traitement syntaxique : concerne liaisons, élisions, homographes et la dénasalisation
 - traitement des liaisons et de la dénasalisation
 - traitement des pauses
 - traitement des élisions
- b) - Buts : - synthèse de la parole à partir de textes écrits
- c) - Matériel : Inconnu au vu des documents consultés.

6.

"Traduction phonétique de textes écrits en français"

B. Prouts

- a) - Méthode : Traitement du texte à transcrire en 4 phases
- "Les caractères de la chaîne d'entrée sont transcrits suivant un code regroupant les lettres présentant certaines analogies de traitement" (cf. GALF, p. 251)

- conversion nombre - graphèmes
- insertion de marqueurs prosodiques et de marqueurs de liaisons
- transcription à l'aide d'un dictionnaire d'environ 1000 éléments, contenant pour chacun les renseignements suivants :

- longueur de la chaîne graphémique
- indicateur : (- du cas général
 - d'appel à une procédure particulière
 - d'une liaison amont interdite
 - d'un mot non prosodique
 - de la simultanéité des deux cas précédents
- chaîne graphémique
- chaîne phonétique
- adresse de différents pointeurs

- b) - Buts :
- synthèse de la parole à partir d'un texte écrit
 - synthèse en temps réel
 - mise en place automatique de marqueurs prosodiques
 - modification aisée des règles.

- c) - Matériel : Microprocesseur INTEL 8080

7.

Présentation rapide du travail contenu dans ce mémoire

- a) - Méthode : Traitement du texte à transcrire en une phase. Au fur et à mesure de son examen le texte est transcrit en tenant compte :
- de la liaison
 - de l'élision
 - de la dénasalisation
 - de la conversion nombre - graphèmes

La transcription mot à mot est exclue, bien que le texte ne soit examiné qu'une fois, car si un mot n'est suivi d'aucun signe de ponctuation, on tiendra compte de la présence dans le texte du mot qui le suit. L'utilisa-

tion d'une grammaire de transcription adaptée concourt à l'obtention du résultat.

b) - Buts : outil d'expérience demandant :

- souplesse
- commodité d'emploi
- compréhension

c) - Matériel:- mini ordinateur : LSI 11/2

- imprimante : Télétype ASR 33
- console : Decscope VT 55

Le tableau suivant va tenter de situer le travail présenté ici (n° 7), par rapport aux 6 méthodes vues (numéros 1 à 6) :

numéro de la méthode	modifications aisées de la "grammaire"	programme adapté à un utilisateur non informaticien	règles de transcription transparentes	mise en place automatique d'une règle	aide à la correction des erreurs dans la grammaire	statistiques élémentaires	gain en vitesse	gain en place mémoire	transcription en vue immédiate de la synthèse	un seul examen du texte pour la transcription
1	?	?	X	?	?	- 1	1	1	1	- 1
2	- 1	- 1	X	- 1	?	- 1	1	1	1	1
3	?	?	X	?	?	1	?	?	1	- 1
4	1	1	- 1	- 1	?	1	1	1	1	- 1
5	?	?	X	?	?	- 1	- 1	?	1	- 1
6	1	?	- 1	?	?	1	1	1	1	- 1
7	1	1	1	1	1	1	- 1	- 1	- 1	1

gende : 1 = oui

? = réponse inconnue au vu des documents consultés

- 1 = non

X = méthodes incomparables sur ce point

E T U D E

INTRODUCTION

Le travail présenté comporte deux parties distinctes et indépendantes :

- Construction et modifications d'une grammaire de transcription.
- Transcription d'un texte à l'aide de cette grammaire.

Ces deux parties sont dues à un manque d'espace mémoire pour les raisons suivantes :

- le matériel utilisé, un mini-ordinateur, qui limite cet espace,
- l'objectif prioritaire que l'on s'est fixé : la commodité, qui peut augmenter la taille de l'espace mémoire nécessaire.

Si elles sont indépendantes dans leur fonctionnement, elles le sont aussi dans leur utilisation : on verra par la suite, qu'il n'est pas indispensable de passer incessamment de l'une à l'autre.

Dans les deux chapitres suivants va être développée, pour chacune d'elles, la manière dont elles ont été conçues en vue de répondre à cette priorité qu'est la commodité. La réalisation informatique de chacune de ces deux parties sera décrite dans les chapitres suivants.

CHAPITRE II

Construction et modifications de la grammaire de transcription

1°- LA GRAMMAIRE :

La transcription orthographique-phonétique sera faite au moyen d'une grammaire de transcription, c'est-à-dire d'un ensemble de règles de transcription. Une règle de cette grammaire indiquera un lien existant entre graphèmes et phonèmes. Afin de faciliter sa construction et les modifications dont elle pourra être l'objet, on souhaite pour cette grammaire :

- que la formalisation des règles soit transparente à l'utilisateur,
- que les règles soient indépendantes les unes des autres afin d'éviter une compilation de la grammaire après toute modification,
- qu'une règle soit facilement accessible.

Cette grammaire ne contiendra que des règles de transcription orthographique phonétique : par exemple, elle ne sera pas chargée de réaliser une conversion nombre - graphème, ou de localiser une liaison. Ces problèmes seront traités par le programme qui transcrira un texte, et donc on n'y fera pas allusion dans cette première partie de l'étude, consacrée à la grammaire de transcription.

1- 1) Conception et formalisation des règles.

La grammaire de transcription est une grammaire contextuelle, à base de symboles orthographiques uniquement. Dans l'avant-propos du livre de V.G. GAK, Nina Catach a écrit : "Ainsi, grâce à nos 26 lettres, nous parvenons à combiner plus de 130 groupes ou graphèmes différents, qui peuvent atteindre plus de 300 valeurs suivant le contexte" (V.G.Gak, p. 7 rq. 3).

Ceci peut justifier l'emploi d'une grammaire contextuelle. D'autre part, on peut trouver dans ce même ouvrage (p.25-6) de V.G.GAK, les différents cas exprimant "le désaccord entre le son et le signe", et que doit distinguer la grammaire de transcription :

- une lettre correspondant à plusieurs phonèmes : aXe
 - plusieurs lettres correspondant à un seul phonème, pAIN
 - une lettre correspondant à des phonèmes différents, suivant le contexte où elle est placée : aCaCia
 - une lettre muette : pont
- en plus du cas régulier où une lettre correspond à un phonème donné.

Donc les éléments qui doivent figurer dans une règle appartenant à une grammaire contextuelle, sont les suivants :

- le (ou les) caractère (s) transcrit (s)
- les contextes droit et gauche
- le (ou les) phonème (s) correspondant (s)

Afin de symboliser le mieux possible la situation des éléments en question dans une chaîne graphique et leur rôle dans la transcription une règle sera formalisée de la manière suivante :

$$\text{(contexte gauche) + (caractère (s) transcrit (s) + (contexte droit) = [phonèmes]}$$
$$\text{soit } C G \quad + \quad C T \quad + \quad C D \quad = \quad [P H]$$

1-2) Symboles pouvant figurer dans le texte d'une règle.

On peut les regrouper suivant trois catégories : ceux figurant dans

- 1 - le (ou les) caractère (s) transcrit (s) (encore désigné par C T ou terme médian).
- 2 - les contextes (C G et C D)
- 3 - le (ou les) phonèmes (s) (P H)

En fin de ce paragraphe figure une liste d'exemples numérotés, correspondant aux exemples mentionnés dans le texte.

1 - 2 - 1) Symboles pouvant figurer dans le terme médian.

Lettres alphabétiques :

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Symboles diacritiques :

- 1 (= accent aigu)
- 2 (= accent grave)
- 3 (= accent circonflexe)
- 4 (= tréma)
- 5 (= cédille)

- Une ou plusieurs lettres alphabétiques, suivies ou non de symboles diacritiques peuvent figurer dans le terme médian (ex. 1-1). Cette symbolisation des signes diacritiques provient de l'absence de ceux-ci sur la console et l'imprimante utilisées. D'autre part, un symbole diacritique est toujours le suivant immédiat du caractère qu'il modifie.

- Le symbole "␣" ou espace, ou encore "blanc" est accepté s'il est ^{l'unique} caractère transcrit dans la règle. Son rôle est d'indiquer que cette règle concerne les liaisons (ex. 1-2).

1 - 2 - 2) Symboles pouvant figurer dans les contextes gauches et droits.

Tout d'abord, il faut signaler que ceux-ci sont limités quant à leur nombre d'éléments (l'explication en sera donnée au chapitre 4). On peut les classer suivant quatre catégories :

- a) Lettres alphabétiques suivies ou non de diacritiques. (ex. 2 - 1)
- b) "classes"

Une classe est un ensemble choisi de caractères alphabétiques. Elles sont au nombre de 9 :

	nom	définition en compréhension	définition en extension
1	CONSO	consonnes	B C D F G H J K L M N P Q R S T V W X Z
2	LIKID	consonnes "liquides"	L R
3	CNONL	consonnes non "liquides"	B C D F G H J K M N P Q S T V W X
4	NASAL	consonnes "nasales"	N M
5	CNONN	consonnes non "nasales"	B C D F G H J K L P Q R S T V W X
5	VOYEL	voyelles	A E I O U Y
7	E-Y-I	e, i, y	E I Y
8	A-O-U	a, o, u	A O U
9	A ___ Z	l'alphabet	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

remarques :

- le ç, noté C5, est considéré comme une consonne codée sur deux caractères, dans la classe "CONSO".

Toute voyelle suivie d'un symbole diacritique est considéré comme une voyelle codée sur deux caractères dans la classe VOYEL (ex. 2-2)

- Dans une règle, le nom d'une classe apparaît entre guillemets, obligatoirement. (ex. 2-3)

Raisons du choix de ces classes.

- Sauf exception (notées ci-dessus) elles sont formées d'éléments codés sur un seul caractère, contrairement aux "classes" définies par Divay & Guyomard (thèse p. 48 § III) qui contiennent des éléments de longueur variable. La justification en sera donnée dans le paragraphe concernant le classement des règles (chap. II - § 2.2).

- Ces classes ne peuvent être modifiées sans le changement de certaines instructions du programme-source :
le fait de pouvoir en modifier le contenu a été jugé sans trop d'intérêt pour les raisons suivantes :

a) - encombrement de la mémoire de l'utilisateur, ou des documents qu'il devra consulter, lors de corrections dans la grammaire de transcription. Dans le système de transcription de Divay & Guyomard (Divay & Guyomard p. 110-111-112) figurent au total 73 classes.

b) Ces 9 classes donnent satisfaction pour la constitution d'une grammaire de transcription répondant aux objectifs fixés (cf. tableau p. 16).

c) 4 symboles particuliers.

1° " # "

- Il indique une extrémité de mot (notation particulière pour phonéticiens et linguistes) - (ex. 2-4)

2° "&"

- Il doit être précédé d'un nom de "classe"
- Il indique le même élément que celui que l'on considère dans la "classe" en question (ex. 2-5)

3° - "/" (sauf)

- Il doit être suivi d'un caractère alphabétique ou d'un nom de liste.
- Il indique :
 - a) - s'il est suivi d'une voyelle (ex. 2-6) : n'importe quelle voyelle sauf celle mentionnée.
 - b) - s'il est suivi d'une consonne (ex. 2-7) : n'importe quelle consonne sauf celle mentionnée.
 - c) - s'il est suivi d'un nom de classe (ex. 2-8) : n'importe quel caractère (soit lettre alphabétique ou symbole diacritique) non contenu dans la classe mentionnée.

4° - "┘"

- Si un tel symbole figure dans le contexte droit (resp. gauche) il ne peut être suivi (resp. précédé) que par un élément identique ou se trouver à l'extrémité du contexte (ex. 2-9)
- son rôle n'apparaît pas directement dans la transcription d'un texte, mais dans l'agencement de la grammaire (cf. Chap. II - § 2.2.2.).

d) Eléments permettant une syntaxe correcte et non ambiguë de l'écriture des règles : +, (,)

Toute règle doit répondre aux 5 conditions suivantes:

- Un nom de classe doit être séparé de tout élément qui le suit dans le contexte par un signe + (ex. 2-10).
- Le symbole "/" n'est pas séparé de l'unique symbole dont il modifie la valeur (classe ou caractère) mais l'ensemble de ces deux éléments est séparé de l'élément suivant du contexte par un signe + (ex. 2-11)

- Le symbole "&" est séparé de l'élément suivant du contexte par un signe + (ex. 2-12)

- Un caractère alphabétique, un symbole diacritique, ou un des symboles " " et " " est obligatoirement séparé de l'élément du contexte qui le suit par un signe +

si celui-ci est : soit un nom de classe

soit le symbole "/" ou "&"

si non, l'emploi du signe + est facultatif (ex. 2-13)

- Si un signe + (au moins) figure dans le contexte gauche ou droit d'une règle, celui-ci doit être placé entre parenthèse (ex. 2-14).

L'ensemble de ces conditions a pour but :

- de reconnaître les caractères illégaux
- de bien distinguer le "squelette" d'une règle.

remarque : C G ou C D peuvent ne contenir aucun élément, la règle possède alors un contexte gauche ou droit, vide.

1 - 2 - 3) Phonèmes

Les phonèmes pris en compte, sont au nombre de 35.

La console et l'imprimante utilisées étant dépourvues des symboles du code phonétique international, il a été nécessaire de coder ces symboles. Ce codage aurait pu se faire à l'aide d'un caractère par phonème, mais, afin de le rendre un peu plus "perceptible" à la lecture, il a semblé préférable de le réaliser au moyen de deux caractères pour chaque phonème. Pour ce faire, il a été très utile de consulter le "Bulletin du groupe reconnaissance automatique de la parole; AFCET" (vol. 5, N° 2, déc. 76) qui donne les codes phonétiques utilisés pour différents projets. Voici la liste des 36 phonèmes utilisés dans le cas présent, avec leur codage respectif et un exemple orthographique pour chacun d'eux :

PHONEMES

ø	EU	peu
ɛ	AI	geai
œ	OE	seul
ẽ	IN	pain
o	AU	dos
œ	UN	brun
u	OU	sous
ã	AN	lent
õ	ON	bon
j	IE	fille
ʃ	CH	chat
ɲ	GN	vigne
ə	ə	parole
ɪ	I	riz
e	E	blé
y	U	rue
a	A	pas

ɔ	O	sol
ɦ	Y	lui
w	W	roi
s	S	seau
z	Z	rose
ʒ	J	géant
f	F	fourmi
v	V	vent
n	N	nid
m	M	mouche
l	L	lapin
r	R	rat
p	P	poule
b	B	bol
t	T	tas
d	D	dé
g	G	gris
k	K	clé

Une règle peut donner 1, 2 ou 3 phonèmes (ex 3 - 1)

Le phonème ou le groupe de phonèmes figurant dans une règle constitue le second membre de celle-ci, ou de l'équation " qui la représente. Il est encadré de crochets, notation habituelle chez les phonéticiens, permettant de délimiter le nombre exact de caractères qui le constituent.

Exemples

Numéro	REGLE	Exemple d'application	
1-1	+a3+= [A ₁]	châle(cha3le)	CH A L
1-1	+a+= [A ₁]	chalet	CH A L AI
1-1	+ain+= [IN]	pain	P IN
1-2	ilst _L + "vowel" = [Z _L]	ils ont	I L Z ON
2-1	/#te+x= [AI]	exact	AI G Z A
2-2	"vowel"+s+"vowel" = [Z _L]	résidu	R E Z I D U
2-3	+cht+"conso" = [K _L]	chlore	K L O R
2-4	+illl+= [I _L L _L]	illogique	I L O J I K
2-4	+ts+ # = [L _L]	petits	P O E T I
2-5	te+("conso"+&) = [AI]	pelle	P AI L
2-6	gte+/u= [L _L]	geai	J AI
2-7	/s+t+ion= [S _L]	action	A K S IE ON
2-8	e+("conso"+/"vowel") = [AI]	pastel	P A S T AI L
2-9	/#et _L + ("vowel"+L _L) = [L _L]	et aussi	E AU S I
2-10	te+("conso"+#) = [AI]	gel	J AI L
2-11	+cht(/"vowel"+e) = [K _L]	chrétien	K R E T IE IN

2-12	$+e(\text{"conso"}+\&+/) = [AI]$	pressant	P R A I S A N
2-13	$+b(s+!/)= [P_{v_i}]$	absent	A P S A N
2-13	$(\#+\text{"conso"})+e+(s+) = [AI]$	les	L A I
2-13	$(\#+\text{"conso"})+e+s = [AI]$	les	L A I
2-14	$(!+\text{"conso"})+e+\# = [OE]$	le	L O E
3-1	$+x+= [K_{v_i} S_{v_i}]$	axe	A K S

Remarque à propos de ces exemples :

Une règle présentée dans cette liste peut ou non figurer dans une grammaire de transcription.

- si elle y figure, employée seule, elle n'a pas réellement de sens, mais convient au mot choisi (en particulier ex B - 9)

- si elle n'y figure pas, elle a été notée uniquement comme illustration, mais convient au mot choisi (elle est alors dotée d'un stérisque (*)).

De plus, il convient de rappeler qu'une règle de grammaire isolée de la liste dont elle fait partie, perd une partie de sa signification (cf. § 2-2 chap.II)

1 - 3) Analyse syntaxique d'une règle.

Lorsque l'utilisateur décide d'ajouter une règle à la grammaire de transcription utilisée, il devra taper à la console cette règle, formalisée selon les indications précédentes. Pour rendre possible, du point de vue informatique, la construction de la grammaire, les règles incorrectes seront refusées. Mais afin de faciliter la tâche de l'utilisateur, le refus d'une règle lui sera bien sûr indiqué, ainsi que l'endroit de la règle où a été détectée la première erreur et la nature de celle-ci.

Ceci est un des rôles d'un sous programme qui détecte au total 19 erreurs différentes, mais s'arrête à la première erreur trouvée dans une règle.

Voici pour chacune des 19 erreurs différentes, la manière dont elle est indiquée sur la console pour une règle incorrecte, et la règle correcte correspondante :

ERREUR: (dans le cas d'une grammaire de décomposition syllabique)
Z+ +Y#=[Z]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 3
ON N EFFECTUE PAS LES LIAISONS LORS D UNE DECOMPOSITION SYLLABIQUE

ERREUR:
Z+ Y+=[Z]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 4
ERREUR : POUR UNE LIAISON, LE TERME MEDIAN NE COMPORTE
QU UN CARACTERE:LE "BLANC"

REGLE D INDICE: 207
Z+ +Y=[Z]

ERREUR:
+E+(#+"C0NS0")=[0E
1 5 10 15 20 25 30 35
ERREUR C0NSTATEE VERS LE CARACTERE 17
ERREUR P0SSIBLE : IL MANQUE LE DERNIER CARACTERE:]

ERREUR:
#+AMER+#=[A M AIR]
1 5 10 15 20 25 30 35
ERREUR C0NSTATEE VERS LE CARACTERE 17
ERREUR : NE S0NT PRIS EN C0MPTE QUE DES GR0UPES DE TR0IS
PH0NEMES ,AU MAXIMUM

ERREUR: (dans le cas d'une grammaire de transcription)
+C+=[C]
1 5 10 15 20 25 30 35
ERREUR C0NSTATEE VERS LE CARACTERE 6
ERREUR P0SSIBLE : PH0NEME INEXISTANT

REGLE D INDICE: 33
+C+=[K]

ERREUR:
+C4+=[S]
0 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 3
ERREUR POSSIBLE : MAUVAIS EMPLOI DES DIACRITIQUES
REGLE D INDICE: 32
+C5+=[S]

ERREUR:
(#+"CONSØ")+E+("CONSØ"+#)=[AI]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 26
ERREUR POSSIBLE : MAUVAIS PARENTHEPAGE

REGLE D INDICE: 215
(#+"CONSØ")+E+("CONSØ"+#)=[AI]

ERREUR:
/ST+I+"VØYEL"=[S IE]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 2
ERREUR POSSIBLE : (/) NE PEUT ETRE SUIVI QUE PAR UN
GUILLEMET ØU PAR UN UNIQUE CARACTERE

REGLE D INDICE: 109
/S+T+(I+"VØYEL")=[S]

ERREUR:
(#"CONSØ"+E+#=[ØE]
1 5 10 15 20 25 30 35
L INDICE DU PREMIER CARACTERE TRANSCRIT EST SUPERIEUR A 44;
LA REGLE NE PEUT ETRE COMPLETE

REGLE D INDICE: 66
(#"CONSØ")+E+#=[ØE]

ERREUR:
+1E+=[E]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 2
ERREUR POSSIBLE : CARACTERE ILLEGAL

REGLE D INDICE: 68
+E1+=[E]

ERREUR:
+A=[A]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 3
ERREUR POSSIBLE : CARACTERE ILLEGAL ØU ØMISSION DU
TROISIEME TERME DU PREMIER MEMBRE

REGLE D INDICE: 21
+A+=[A]

ERREUR:

"VØYEL"+S+"VØYEL"=[Z]

1 5 10 15 20 25 30 35

ERREUR CØNSTATEE VERS LE CARACTERE 8

ERREUR POSSIBLE: ERREUR DE SYNTAXE

REGLE D INDICE: 137

"VØYEL"+S+"VØYEL"=[Z]

ERREUR:

"VAYEL"+S+"VØYEL"=[Z]

1 5 10 15 20 25 30 35

ERREUR CØNSTATEE VERS LE CARACTERE 2

ERREUR POSSIBLE : NØM DE LISTE INEXISTANT

REGLE D INDICE: 137

"VØYEL"+S+"VØYEL"=[Z]

ERREUR:

+CH+"LIKIDE"=[K]

1 5 10 15 20 25 30 35

ERREUR CØNSTATEE VERS LE CARACTERE 12

ERREUR POSSIBLE: ERREUR DE SYNTAXE

ERREUR:

+CH+"LIQUID"=[K]

1 5 10 15 20 25 30 35

ERREUR CØNSTATEE VERS LE CARACTERE 6

ERREUR POSSIBLE : NØM DE LISTE INEXISTANT

REGLE D INDICE: 33

+CH+"LIKID"=[K]

ERREUR:
+AIN+(#+/!VØYEL")=[IN-]
1 5 10 15 20 25 30 35
ERREUR CØNSTATEE VERS LE CARACTERE 10
ERREUR PØSSIBLE : (/) NE PEUT ETRE SUIVI QUE PAR UN
GUILLEMET ØU PAR UN UNIQUE CARACTERE

REGLE D INDICE: 32
+AIN+(#+/!VØYEL")=[IN]

ERREUR:
+E+(&+"CØNSØ")=[AI]
1 5 10 15 20 25 30 35
ERREUR CØNSTATEE VERS LE CARACTERE 5
ERREUR PØSSIBLE : LE SYMBØLE "&" REPRESENTE LE DERNIER CARAC-
TERE QUI LE PRECEDE

ERREUR:
"VØYEL"+S+S"VØYEL"=[Z]
1 5 10 15 20 25 30 35
ERREUR CØNSTATEE VERS LE CARACTERE 12
ERREUR PØSSIBLE:ERREUR DE SYNTAXE

REGLE D INDICE: 137
"VØYEL"+S+"VØYEL"=[Z]

ERREUR:
(#+"CONSØ")+E("CONSØ")#]=[AI]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 23
ERREUR POSSIBLE : MAUVAIS PARENTHEPAGE

ERREUR:
(#+"CONSØ")+E("CONSØ")+#]=[AI]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 23
ERREUR POSSIBLE : MAUVAIS PARENTHEPAGE

REGLE D INDICE: 215
(#+"CONSØ")+E("CONSØ"+#]=[AI]

ERREUR:
+ØU+=ØU
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 6
LE SIGNE = EST TOUJOURS SUIVI DE (

REGLE D INDICE: 116
+ØU+=[ØU]

ERREUR:
+CHCONSØ+=[K]
1 5 10 15 20 25 30 35
ERREUR CONSTATEE VERS LE CARACTERE 9
ERREUR POSSIBLE : PLUS DE 6 CARACTERES DANS LE TERME CENTRAL

REGLE D INDICE: 28
+CH+"CONSØ"=[K]

Conclusion :

Cette formalisation des règles, ainsi que l'ensemble des symboles qui sont autorisés, favorisent l'"unification" de la grammaire, et donc facilitent les modifications dont elle pourra être l'objet. En effet, elle réunit les règles de transcriptions concernant :

cas considéré	exemple de règle	exemple d'application
1) - mots "ordinaires"	+ ou + = [OU]	bouleau
2) - exceptions	(# + "conso") + er+ = [AIR]	MER
3) - liaisons	# et + + "vowel" = []	riche et avare
4) - élisions	+ e+ (# + "vowel") = []	richE et avare
5) - dénasalisation	+ o+ (N# + "vowel" = [o])	bON appétit

2- FONCTIONNEMENT DE LA GRAMMAIRE :

Le fonctionnement de la grammaire, au cours d'une transcription, a été conçu de manière à le rendre facilement compréhensible par son utilisateur, et favoriser ainsi la correction des erreurs de transcription.

2-1 Validation d'une règle de transcription

- soit une chaîne graphique quelconque, notée
xxxyyyyy

où les symboles x représentent la partie de la chaîne déjà transcrite, et les symboles y la partie de la chaîne restant à transcrire.

- Une règle, quelconque elle aussi, s'écrit :

$$\left(\begin{array}{c} \text{contexte} \\ \text{gauche} \end{array} \right) + \left(\begin{array}{c} \text{caractère (s)} \\ \text{transcrit (s)} \end{array} \right) + \left(\begin{array}{c} \text{contexte} \\ \text{droit} \end{array} \right) = \left[\text{phonèmes} \right]$$

les contextes gauche ou droit pouvant être vides.

Dans ces conditions, une règle pourra s'appliquer à cette chaîne graphémique, si et seulement si il y a identité entre la chaîne et la règle pour les caractères de CT, et identité ou appartenance pour les termes de CG et CD.

Exemples d'application ou de non application de règles :

Chaîne graphémique à transcrire :

A C H A T

x y y y y

(# indique l'extrémité d'un mot)

* Soit les règles :

(1) :	+ ai +	=	[AI]
(2) :	+ ch + "vowel"	=	[CH]
(3) :	+ c +	=	[K ₁]
(4) :	+ cc +	=	[K ₂]
(5) :	+ ch + "conso"	=	[K ₃]

Dans ce cas deux règles distinctes peuvent s'appliquer :

les règles (2) et (3).

Il faut donc lever cette ambiguïté.

2.2. Partition et classement des règles d'une grammaire de transcription

A propos de l'exemple présenté ci-dessus, on peut faire deux remarques.

- La règle 1 est rejetée d'emblée.
- Deux règles différentes sont applicables dans les mêmes conditions.

La partition des règles va éviter l'examen de règles qui sont rejetées immédiatement.

Le classement des règles va conduire à une unique possibilité de résultat et aussi à une diminution du nombre de règles considérées.

2.2.1. Partition des règles .

Elle va s'effectuer en 27 sous-ensembles .

Il faut ici se rappeler (cf.chap.2 § 1.2.1.) que seuls 27 caractères peuvent apparaître en première position dans le terme médian d'une règle. Ce sont : A, B Z , et le "blanc".

C'est d'après le premier caractère de son terme médian qu'une règle va être affectée à l'un des 27 sous-ensembles de la partition.

Sont maintenant séparés :

- le sous-ensemble des règles concernant les liaisons.
- chacun des 26 sous-ensembles regroupant les règles dont le premier caractère transcrit est l'une des 26 lettres de l'alphabet.

Il en résulte que pour transcrire une chaîne graphémique débutant par une caractère quelconque, ou pour effectuer une liaison, il suffira d'examiner les règles de transcription appartenant au sous-ensemble adéquat de la grammaire.

2.2.2. Classement des règles.

La partition des règles, si elle permet un gain de temps, n'élimine pas pour autant la possibilité d'un résultat multiple. Dans l'exemple proposé au paragraphe 2-1, on se bornait à l'examen de l'ensemble des règles 2,3,4 et 5, mais le résultat serait le même. Les règles 2 et 3 seraient validées.

Les deux remarques suivantes peuvent conduire au mode de classement des règles, ainsi qu'à l'ordre d'examen de celles-ci :

- Toute règle de la grammaire doit pouvoir être utilisée.
- Si une règle est validée dans un cas particulier, toute règle incluse dans la précédente le sera automatiquement.

a) Explication des termes employés.

1°-

Une règle A est incluse dans une règle B si :

la suite des éléments de CT U CD de A est identique au début de la suite des mêmes éléments dans B, et si la suite des éléments du CG de A (pris de droite à gauche) est identique au début de la suite des mêmes éléments de B.

ex: A1 : + S + = [S□]
 B1 : "vowel" + S + "vowel" = [Z□]

 A2 : + C + = [K□]
 B2 : + CH + = [CH]

2°.

"Longueur droite" d'une règle, ou LD :

LD compte le nombre de caractères transcrits auxquels on ajoute le nombre d'éléments figurant dans CD.

ex: + S + = [S□] LD = 1
 "vowel" + S + "vowel" = [Z□] LD = 2

3°.

"Longueur gauche" d'une règle en LG :

LG compte le nombre d'éléments figurant dans CG

ex : + S + = [S□] LG = 0
 " vowel" + S + "vowel" = [Z□] LG = 1

Ainsi, si la règle A est incluse dans la règle B, nécessairement :

LD (A) \leq LD (B) et LG (A) \leq LG (B)

On a pour les exemples cités en 1°

$$LD (A1) = 1 \leq LD (B1) = 2 \quad \text{et} \quad LG (A1) = 0 \leq LG (B1) = 1$$

$$LD (A2) = 1 \leq LD (B2) = 2 \quad \text{et} \quad LG (A2) = 0 \leq LG (B2) = 0$$

b) Justification de la priorité donnée dans le classement au contexte droit par rapport au contexte gauche.

Les règles de transcriptions utilisées ici présentent peu souvent un contexte gauche non vide, si elles concernent l'une des 26 lettres de l'alphabet. Par contre, il en va tout autrement pour les règles concernant les liaisons. Mais, au vu du nombre total des règles présentant un contexte gauche non vide, et de la possibilité qui est laissée à l'utilisateur de modifier comme il l'entend le classement automatique de celles-ci, priorité est donnée au contexte droit par rapport au contexte gauche, c'est-à-dire que les règles sont d'abord classées dans un certain ordre dépendant de LD, et à LD égales, dans un ordre relatif à LG.

c) Classement des règles.

L'ensemble de la grammaire a été partitionné en 27 sous-ensembles de règles. Chacun d'eux sera classé indépendamment des 26 autres. Dans un sous-ensemble les règles sont ordonnées suivant LD décroissant (au sens large), et, à LD égales, suivant LG décroissant (au sens large).

Le problème qui se poserait dans le classement des règles suivantes :

$$A + B + = [XX]$$

$$+ B + C = [YV]$$

si l'on avait à transcrire la chaîne graphémique ABC, est écarté.

En effet, on suppose qu'il n'y a pas de règles ambiguës dans la grammaire, c'est-à-dire de règles dont l'application, dans un cas bien particulier, donnerait un résultat tributaire de leur ordre de rangement dans un sous-ensemble. Le problème ne se poserait pas dans l'exemple ci-dessus, s'il y avait trois règles au lieu de deux :

$$\begin{aligned} A + B + C &= [ZZ] \\ + B + C &= [YY] \\ A + B + &= [XX] \end{aligned}$$

D'ailleurs ce sont l'habitude et la culture qui, dans le cas de deux règles, déterminent leur bonne utilisation. Mais, comment peut-on faire passer sur machine ces deux sources de savoir, propriétés des êtres humains ?

exemple de classement :

lettre A

```
REGLE D INDICE: 16
F+AI+(S+"VØYEL")=[ØE]
REGLE D INDICE: 9
+AI1+#=[IN]
REGLE D INDICE: 8
+AIN+#=[IN]
REGLE D INDICE: 6
+AIN+"CNØNN"=[IN]
REGLE D INDICE: 134
+A+(I+L+L)=[A ]
REGLE D INDICE: 139
+A+(I+L+#)=[A ]
REGLE D INDICE: 4
P+AY+S=[AI ]
REGLE D INDICE: 13
+AN+#=[AN]
REGLE D INDICE: 12
+AI3+=[AI]
REGLE D INDICE: 11
+AN+"CNØNN"=[AN]
REGLE D INDICE: 10
+AN+"CNØNN"=[AN]
REGLE D INDICE: 163
+A+(Y+"VØYEL")=[AI]
REGLE D INDICE: 20
+AI+=[AI]
REGLE D INDICE: 19
+AU+=[AU]
REGLE D INDICE: 18
+A3+=[A ]
REGLE D INDICE: 17
+A2+=[A ]
REGLE D INDICE: 21
+A+=[A ]
```

Bien que ce classement soit fait de façon automatique lors de l'insertion de toute règle dans la grammaire, l'utilisateur dispose, par la suite, des éléments nécessaires pour placer où il l'entend la règle de son choix sans pour autant fausser le classement dans la grammaire des règles qui y seront ajoutées.

- Deux règles de longueurs identiques (LG et LD) peuvent être inversées sans difficulté, si elles appartiennent au même sous-ensemble de la grammaire.

- Deux règles quelconques peuvent être amenées à avoir des longueurs LG et LD égales, en ajoutant dans les contextes gauche ou droit, des symboles "⌊" en quantité suffisante (cf. 1-2-2-c .), ce qui ne modifie en aucune manière, le rôle de la règle en question.

Ce classement des règles, non automatique, est nécessaire pour bien positionner deux règles de longueurs identiques, si l'une des deux est un cas particulier de l'autre. Ceci est différent de la notion d'inclusion définie en a) , qui se base sur l'identité des symboles représentant les éléments des contextes.

ex. 1	+ E + ("conso" + ⌊)	= [AI]	martinet, sel, nef
2	+ E + Z #	= [E⌊]	nez, avez..

Dans l'exemple ci-dessus, la règle n° 2, tout en n'étant pas incluse dans la règle n° 1 (écriture différente), est un cas

particulier de celle-ci. Il est donc nécessaire, si l'on désire pouvoir appliquer la règle n° 2, de la placer avant la règle n° 1, dans le sous ensemble de règles correspondant au caractère E.

- D'autre part, le classement automatique dans la grammaire de toute règle nouvellement insérée, permet de contrôler s'il n'existe pas déjà une règle identique, auquel cas une erreur est signalée à la console.

d) Méthode d'examen des règles.

Soit Y le premier caractère non transcrit de la chaîne graphique. On parcourt les règles du sous-ensemble correspondant dans leur ordre de classement (automatique ou non automatique). La recherche d'une règle applicable s'arrêtera à la première règle trouvée. Si aucune règle ne peut s'appliquer dans ce cas, une erreur est signalée à l'utilisateur.

2-3) Méthode de transcription d'un mot isolé.

Pour achever la description de la méthode, couramment utilisée, il manque, en plus des explications données en 2-1 et 2-2, un élément. Il s'agit de la progression dans le mot à transcrire. Lorsqu'une règle est validée, on avance d'un nombre de caractères égal au nombre de caractères transcrits par cette règle.

Comment s'effectue donc la transcription d'un mot ?

- 1) positionnement en début de mot.
- 2) Examen des règles de la liste classée correspondant au caractère examiné.
- 3) Application de la première règle validée.
- 4) Progression dans le mot.

5) Si transcription du mot non achevée, retour au 2

Fin.

Ceci peut être vérifié sur l'exemple suivant, obtenu sur

listing :

REGLE D INDICE: 91
+Ø+(I+N+"CNØNN")=[W]
REGLE D INDICE: 92
+Ø+(I+N+#)=[W]
REGLE D INDICE: 108
+ØM+"CNØNN"=[ØN]
REGLE D INDICE: 110
+ØN+"CNØNN"=[ØN]
REGLE D INDICE: 112
+ØEU+=[ØE]
REGLE D INDICE: 88
+Ø+(S+#)=[AU]
REGLE D INDICE: 101
+Ø+(Y+"VØYEL")=[W A]
REGLE D INDICE: 125
+ØU2+=[ØU]
REGLE D INDICE: 185
+ØU3+=[ØU]
REGLE D INDICE: 200
+Ø+(S+"VØYEL")=[AU]
REGLE D INDICE: 201
+Ø+(N+H)=[Ø]
REGLE D INDICE: 32
+ØN+#=[ØN]
REGLE D INDICE: 115
+Ø3+=[AU]
REGLE D INDICE: 116
+ØU+=[ØU]
REGLE D INDICE: 225
+ØI+=[W A]
REGLE D INDICE: 118
+Ø+=[Ø]

REGLE D INDICE: 103
+MM+=[M]
REGLE D INDICE: 104
+M+=[M]

REGLE D INDICE: 203
+S+(C+"E-I-Y")=[]
REGLE D INDICE: 204
+S+(C+H)=[]
REGLE D INDICE: 137
"VØYEL"+S+"VØYEL"=[Z]
REGLE D INDICE: 175
#+S+#=[S]
REGLE D INDICE: 138
+SH+=[CH]
REGLE D INDICE: 140
+SS+=[S]
REGLE D INDICE: 141
+S+#=[]
REGLE D INDICE: 142
+S+=[S]

REGLE D INDICE: 104
+M+=[H]
REGLE D INDICE: 225
+ØI+=[W A]
REGLE D INDICE: 140
+SS+=[S]
REGLE D INDICE: 32
+ØN+#=[ØN]

TRANSCRIPTION DE : MOÏSSON
M U A S ØN

On peut remarquer que la mise bout à bout des caractères transcrits de chacune des règles, dans leur ordre d'application redonne exactement le mot à transcrire.

3- Corrections des erreurs.

Etant donné que ce travail est supposé convenir à la création de diverses grammaires de transcription, il semble important d'aider à la corrections d'erreurs pouvant survenir lors de ces opérations.

Ceci suppose ;

- se rendre compte de l'erreur,
- localiser, dans le tableau, son origine,
- corriger la grammaire,
- vérifier l'exactitude de la correction.

3-1 Signalisation d'une erreur.

L'erreur est signalée dans un cas uniquement :

- si la transcription ne peut se poursuivre car il n'existe dans la grammaire aucune règle qui puisse s'appliquer dans ce cas particulier.

Le résultat apparaît alors comme une suite de deux paires d'astérisques * * * *

Sinon, il appartient à l'utilisateur de repérer lui-même les autres erreurs qui correspondent alors à une mauvaise transcription.

3-2) Localisation dans la grammaire de la provenance d'une erreur.

On peut à tout moment obtenir la transcription d'un mot, accompagné de la trace des opérations effectuées avant la donnée du mot transcrit, sur la console ou à l'imprimante (cf. tableau p), c'est-à-dire la suite des règles appliquées dans ce cas particulier, cette suite s'arrêtant si le mot entier est transcrit, ou si aucune règle ne peut plus s'appliquer.

Exemple de transcription du mot "maçon" à partir de trois grammaires différentes.

Exemple 1

REGLE D INDICE: 104
+M+=[M]
REGLE D INDICE: 21
+A+=[A]

TRANSCRIPTION DE : MAC5ØN
** **

Exemple 2

REGLE D INDICE: 104
+M+=[M]
REGLE D INDICE: 21
+A+=[A]
REGLE D INDICE: 32
+C+=[K]

TRANSCRIPTION DE : MAC5ØN
** **

Exemple 3

REGLE D INDICE: 104
+M+=[M]
REGLE D INDICE: 21
+A+=[A]
REGLE D INDICE: 33
+C5+=[S]
REGLE D INDICE: 118
+Ø+=[Ø]
REGLE D INDICE: 220
+N+=[N]

TRANSCRIPTION DE : MAC5ØN
M A S Ø N

remarque: On peut noter que toute règle de la grammaire se reconnaît dans l'ordinateur grâce à son "indice", dont la signification sera donnée dans un chapitre concernant la réalisation informatique de ce projet (cf. p.) et qui accompagne automatiquement le texte de toute règle. Ceci favorise l'accessibilité des règles de la grammaire.

* Ceci permet de situer l'erreur dans le mot à transcrire.

On peut reprendre à ce propos, les trois cas de l'exemple précédent :

- maC5on
- maC5on
- mac5On

* A partir de là, il est possible de localiser l'erreur dans la grammaire de transcription : en effet, on peut facilement obtenir la liste classée des règles appartenant à l'une des 27 parties de la grammaire (cf. p.45). Or, puisque l'on sait que la première règle validée dans une telle liste s'applique, on peut en déduire l'absence, le mauvais positionnement, ou la mauvaise formalisation d'une règle, dans le sous-ensemble de la grammaire correspondant au caractère, du mot à transcrire, conduisant à l'erreur.

Reprenons les trois cas de l'exemple précédent :

1 - ex :

Il n'y a aucune règle correspondant au graphème C qui puisse s'appliquer dans ce cas particulier. Il faut introduire dans la grammaire la règle suivante :

+ C 5 + = [S_u], pour transcrire le mot "maçon".

2- ex:

Il n'y a aucune règle correspondant au graphème C5, mais une

autre règle (incluse dans celle-ci) a pu s'appliquer en donnant un mauvais phonème. Il faut donc introduire dans la grammaire la règle suivante :

$$+ C 5 + = [S \sqcup]$$

3- ex:

La règle (+ ON + "C NONN" = [ON]) que l'on pensait pouvoir s'appliquer, ne peut en fait le faire, le symbole n'étant pas contenu dans l'ensemble de caractères "CNONN" (cf. p. chp. II). Il faut introduire dans la grammaire, la règle suivante :

$$+ ON + \# = [ON]$$

3-3 Résultats après correction.

REGLE D INDICE: 104

+M+=[M]

REGLE D INDICE: 21

+A+=[A]

REGLE D INDICE: 28

+C5+=[S]

REGLE D INDICE: 241

+ON+#=[ON]

TRANSCRIPTION DE : MACSON
M A S ON

4- Modifications et renseignements.

Précédemment on a vu qu'il était nécessaire de modifier la grammaire pour sa construction et sa mise au point, et pour ce faire, qu'il était souhaitable de posséder des renseignements la concernant. La grammaire est donc un interpréteur afin de pouvoir faire "retour-arrière" sans trop de difficultés. (Il ne faut pas oublier que la vitesse d'exécution n'est pas le souci majeur de ce projet). De plus, cette grammaire peut être éditée.

Voici tout d'abord les différentes modifications possibles dans toute grammaire, résumées dans un tableau pouvant apparaître à la console selon le désir de l'utilisateur :

MODIFICATIONS

QUESTIONS POSSIBLES	IL VOUS FAUT	ORDRE
I	I	I
I SUPPRESSION D UNE REGLE	I INDICE	I 1 I
I-----I	I-----I	I-----I
I	I	I
I AJOUT D UNE REGLE	I TEXTE	I 2 I
I-----I	I-----I	I-----I
I	I	I
IMODIFICATION=SUPPRESSION+AJOUTI	INDICE+TEXTE	I 3 I
I-----I	I-----I	I-----I
I	I	I
I REGLE D INDICE I A PLACER	I I	I 4 I
I AVANT LA REGLE D INDICE K	I K	I I
I-----I	I-----I	I-----I
I	I	I
I REGLE D INDICE I A PLACER	I I	I 5 I
I APRES LA REGLE D INDICE K	I K	I I
I-----I	I-----I	I-----I

Voici maintenant, toujours résumés sous forme de deux tableaux pouvant apparaître à la console au gré de l'utilisateur, les renseignements qu'il peut obtenir dans la grammaire de transcription utilisée. Ceux-ci peuvent être utiles à des fins différentes :

- corrections d'erreurs dans la grammaire,
- exploitation de statistiques élémentaires obtenues après transcription d'un texte (cf. §.2, ch. III),
- vérification du codage des règles (cf. § 2, ch. IV).

RENSEIGNEMENTS

QUESTIONS POSSIBLES	IL VOUS FAUT	ORDRE
I	I	I I
I LISTE CLASSEE DES REGLES POUR I	I LE	I 6 I
I UN CARAC. OU CAR. PAR CAR.(#)I	I CARACTERE	I I
I-----I	I-----I	I-----I
I	I	I I
I LISTE DES REGLES COMPTANT M	I M	I 7 I
I CARACTERES DS LE TERME MEDIAN I	I	I I
I POUR UN CARACTERE DONNE OU	I CARACTERE	I I
I CARACTERE PAR CARACTERE	I	I I
I-----I	I-----I	I-----I
I	I	I I
I TRANSCRIPTION D UN MOT , TEXTEI	I MOT	I 8 I
I ET INDICE DES REGLES UTILISEESI	I	I I
I-----I	I-----I	I-----I
I	I	I I
I TEXTE D UNE REGLE	I INDICE	I 9 I
I-----I	I-----I	I-----I

QUESTIONS POSSIBLES	IL VOUS FAUT	ORDRE
I	I	I I
I REGLE DONT LE CONTEXTE DROIT	I	I I
I (GAUCHE) COMPORTE M TERMES ,	I CARACTERE	I 10(D)I
I POUR UN CARACTERE DONNE OU	I M	I 11(G)I
I CARACTERE PAR CARACTERE	I	I I
I-----I	I-----I	I-----I
I CODAGE NUMERIQUE D UNE REGLE	I INDICE	I 12 I
I-----I	I-----I	I-----I
I REGLES DONNANT MEME PHONEME	I PHONEME	I 13 I
I-----I	I-----I	I-----I
I REGLES DONNEES DANS L ORDRE	I RIEN	I 14 I
I DES INDICES CROISSANTS	I	I I
I-----I	I-----I	I-----I

Conclusion.

Cette première partie, relative à la mise au point d'une grammaire de transcription a été conçue comme un véritable outil de travail. On l'a voulu d'un emploi simple et commode, mais laissant cependant une grande part de liberté à son utilisateur, non pas dans la forme de ce qu'il construit, mais dans son utilisation future. En annexe on trouvera le détail du dialogue utilisateur - console qui la sous-tend.

CHAPITRE III

Introduction

Cette seconde partie du travail est un complément à la première partie : elle permet l'exploitation d'une grammaire de transcription construite et mise au point dans cette première partie. Son objectif est double :

- 1) Transcription d'un texte à l'aide d'une grammaire quelconque
- 2) Obtention de statistiques élémentaires sur l'emploi des règles de cette grammaire.

D'autre part, pour effectuer la transcription d'un texte, les diverses règles contenues dans la grammaire employée devront suffire. Le programme, correspondant à cette seconde partie et effectuant la transcription du texte au moyen d'une grammaire, devra donc gérer lui-même les cas de :

- ponctuation
- liaison (en partie)
- conversion nombre - graphèmes.

1 - Transcription d'un texte

1.1. Création du fichier contenant le texte à transcrire

a) - Caractères admis dans un tel fichier

Si le texte à transcrire est encore inexistant en mémoire, il faut tout d'abord l'y insérer, ceci est facultatif, car il est tenu compte du cas où le texte figure déjà en mémoire. L'introduction du texte se fait à la console. Tous les caractères, existant sur le clavier de cette dernière, sont acceptés, y compris ceux qui par la suite seront à l'origine

d'erreurs car illégaux. Cependant, les caractères légaux, c'est-à-dire ne pouvant provoquer d'erreurs dues à leur nature, sont les suivants :

x lettres alphabétiques : A B C D E F G H I J K L M N O P Q R S T U V
W X Y Z

x chiffres : 0 1 2 3 4 5 6 7 8 9

x signes de ponctuation (ne sont sujets à aucune phonétisation) :

. , : ; ! ? - () + = [] " / &

(convention : le symbole '(apostrophe) est représenté par le symbole ")

.le symbole # : ne fait pas partie de l'ensemble de caractères à transcrire. Seuls sont pris en compte les caractères qui le précèdent (sur une même ligne). Les suivants sont ignorés.

.D'autre part, il n'existe pas de majuscules et de minuscules; en effet, les caractères utilisés sont uniformes sur la console, et apparaissent sous forme de majuscules.

b) - Mode d'introduction d'un texte

Le texte est introduit ligne par ligne. La longueur maximum d'une ligne est à chaque fois indiquée. Dans tous les cas, il est nécessaire de terminer une ligne par le symbole # : ceci pour éviter la troncation ou l'oubli de mots. Si ce symbole n'est pas rencontré lors de la vérification, il sera demandé à l'utilisateur de réécrire correctement cette ligne. De plus, étant donné que seuls figureront dans le fichier contenant le texte, les caractères précédant ce symbole, il faut porter une attention particulière aux espaces séparant deux mots consécutifs, dont le premier se trouve en fin d'une ligne, et le second en début de la ligne suivante. Enfin, pour indiquer qu'une ligne est la dernière ligne de texte à introduire, il suffit de terminer cette ligne par deux symboles ## consécutifs, au lieu d'un seul. Le fichier contenant le texte est alors clos, et prêt pour la transcription. Voici l'exemple d'un texte que l'on désire transcrire :

- texte correct avant écriture à la console (a)
- écriture du même texte à la console, avec quelques erreurs (b)
- texte tel qu'il figure dans le fichier à transcrire (c)

(texte : "Le vieil homme et la mer" - Hemingway).

a) "Il n'avait pas besoin de boussole pour savoir où était le Sud-Ouest. Il lui suffisait de sentir l'alizé souffler, et la voile virer. Je ferais bien de poser une petite ligne avec une cuiller, histoire d'attraper quelque chose qui me mette un peu de jus dans le corps".

b) - C = console et U = utilisateur

C : (24) Longueur maximale (24)

U : Il n'avait pas besoin de boussole pour

C : (24) Longueur maximale (24)

U : savoir où était le sud-ouest. Il lui suffisait

C : (24) Longueur maximale (24)

U : de sentir l'alizé souffler et la voile virer. Je
fe

C : (24) Longueur maximale (24)

U : rais bien de poser une petite

C : (24) Longueur maximale (24)

U : ligne avec une cuiller, histoire d'attraper quelque
chose qui

C : (24) Longueur maximale (24)

U : me mette un peu de jus dans le corps

c) - Texte, tel qu'il a été enregistré (code orthographique usuel pour les accents et les coupures de mot) :

- Il n'avait pas besoin de boussole pour savoir où était le sud-ouest.
Il lui suffisait de sentir l'alizé souffler et la voile virer.
Je ferais bien de poser une petite ligne avec une cuiller,
histoire d'attraper quelque chose qui me mette un peu de jus dans le corps.

1.2. Choix de la grammaire de transcription

Le texte à transcrire ayant été créé ou choisi parmi les fichiers déjà existants, il reste à déterminer la grammaire à utiliser. Toute grammaire construite dans la première partie convient. Or celles-ci une fois créées, figurent en mémoire (dans la limite de la place disponible), et sont accessibles grâce au nom qui leur a été attribué. Le choix de l'une d'entre elles dépend principalement des résultats que l'on veut obtenir.

Ce choix s'effectue en répondant par le nom de la grammaire choisie, que l'on tape au clavier de la console, à une question apparaissant à la même console, et provenant du programme source :

C : nom du fichier où se trouve la grammaire à utiliser

U : xxxxxx

1.3. Transcription

Le texte à transcrire ayant été choisi, ainsi que la grammaire à utiliser, la transcription peut s'effectuer, sans aucun recours à l'utilisateur. Le texte transcrit sera automatiquement imprimé sur listing.

1.3.1. Méthode de transcription d'un texte

On souhaite que la transcription se fasse, de la manière la plus complète possible, en un seul examen du texte à transcrire. Cette transcription ne devra pas être une simple concaténation des mots isolés transcrits, dans l'ordre de leur apparition dans le texte. La solution retenue est la suivante :

- examiner :
- Soit deux mots consécutifs, s'ils ne sont pas séparés par un signe de ponctuation quelconque (hormis le trait d'union). Dans ce cas, seul le premier mot sera transcrit.
 - Soit un seul mot, s'il est suivi d'un signe de ponctuation différent du trait d'union.

Cette méthode d'une part évite l'assimilation des mots du texte aux mêmes mots prononcés isolément, et d'autre part est en accord avec les idées que Schane a exprimées dans sa thèse (cf. "French Phonology and Morphology", 1977, M.I.T. Press), selon lesquelles la prononciation d'un mot, dans un ensemble de mots, ne dépend pas des mots qui le précèdent, mais dépend uniquement du mot qui le suit (phénomènes de liaison, d'élision, de dénasalisation). Bien entendu, cela ne s'applique pas directement au cas de nombres. Tout d'abord, lors de leur rencontre dans le texte à transcrire, ceux-ci sont orthographiés et donc transformés en un ou plusieurs mots jouant exactement pour la méthode de transcription, le même rôle que les autres mots du texte.

D'autre part, lorsque deux mots consécutifs sont examinés, les problèmes de liaison, d'élision et dénasalisation sont pris en considération. Ceci se révèle impossible, et d'ailleurs sans aucun intérêt dans l'autre cas.

1.3.2. Définition de mot et de nombre

Le texte à transcrire est réduit, pour le programme qui va le traiter, à un ensemble de caractères, dans lesquels il ne peut, à priori, distinguer ni les mots, ni les nombres. Les critères permettant de les extraire d'un ensemble de caractères quelconques, seront aussi utiles à l'utilisateur afin de bien écrire le texte ou encore pour corriger certaines erreurs dans la transcription.

mot : toute suite de caractères alphabétiques ou numériques, consécutifs dans le texte, débutant obligatoirement par un caractère alphabétique

nombre : toute suite de caractères numériques immédiatement précédée par un caractère non alphabétique

exemple : l'an passé, ils étaient 24

Cette phrase s'écrit comme suit dans le fichier à transcrire :

l'an passel, ils eltaient 24

de cette phrase vont être extraits :

5 mots : - l

- an

- passe l

- ils

- éltaient

1 nombre : - 24

Il est à noter que cette manière de différencier mots et nombres, est en accord avec l'emplacement des symboles diacritiques (1, 2, 3, 4, 5) dans le mot.

1.3.3. Transcription d'un mot

Paradoxalement, bien qu'une transcription successive de mots isolés soit exclue, un seul mot est transcrit à la fois. Mais, si l'on se souvient des différents types de règles que peut contenir la grammaire, et du mode d'examen des mots du texte, on s'aperçoit qu'il est possible d'adjoindre le mot suivant celui qui est transcrit, ou les premiers caractères qui le composent, au contexte droit de toute règle, en respectant toutefois le nombre d'éléments autorisés. Par exemple, il est possible de construire des règles répondant à la prononciation des deux expressions suivantes :

- 1) un arbre immobile
- 2) un arbre tourmenté

Bien qu'appartenant à un même mot (arbre), les lettres e ne se prononcent pas de la même manière dans les deux cas :

- 1) élision du e final du mot "arbre"
- 2) le e final de "arbre" se prononce

Ce cas particulier serait résolu par la présence dans la grammaire de transcription des deux règles suivantes :

$$+ e + (\# + \text{"vowel"}) = [_ _]$$
$$+ e + = [OE]$$

UN N A R B R I M Ø B I L UN N A R B R Ø E
T Ø U R M A N T E

1.3.4. Transcription des nombres

a) Généralités

La grammaire de transcription utilisée ne comportant que des règles qui puissent transformer des graphèmes en phonèmes, et d'autre part, désirant que la transcription d'un texte quelconque se fasse au cours d'un seul examen de celui-ci, sans recours à l'utilisateur, il s'avère nécessaire de faire orthographier un nombre par le programme, lors de sa rencontre dans le texte.

b) Insertion de la transformation nombre - graphèmes

Lorsque le pointeur approprié parcourt le texte à transcrire, il reconnaît la présence d'un nombre au premier caractère qui le constitue (cf. 1 - 3 - 2). A partir de ce moment-là, après lecture de tous les caractères numé-

riques constituant le nombre en question, les mots à transcrire ne proviennent plus du texte jusqu'à transcription complète du nombre. Ce nombre subit en effet un traitement spécial qui le transforme en un ou plusieurs mots, constituant la source provisoire des mots à transcrire. Ceci se fait en respectant le mode général de transcription décrit en 1.3.1.

c) Méthode de transformation nombre - graphèmes

Hypothèse de départ : l'orthographe des nombres obéit à des règles de la grammaire française jugées immuables. Cette hypothèse permet de rendre "immuables" les éléments du programme contribuant à cette transformation. Aussi, ceux-ci ne vont pas apparaître sous forme de règles facilement accessibles et modifiables, mais seront partie intégrante du programme, et donc inaccessibles à l'utilisateur.

Ces éléments sont un ensemble des 35 mots ou groupes de mots, nécessaires à la prononciation de nombres entiers inférieurs strictement à 10^{12} . En voici la liste :

Indice	Nombre	mot ou groupe de mots	Indice	Nombre	mot ou groupe de mots
1	1	un	19	19	dix-neuf
2	2	deux	20	20	vingt
3	3	trois	21	30	trente
4	4	quatre	22	40	quarante
5	5	cinq	23	50	cinquante
6	6	six	24	60	soixante
7	7	sept	25	80	quatre-vingt
8	8	huit	26	100	cent
9	9	neuf	27	1000	mille
10	10	dix	28	10^6	million
11	11	onze	29	10^9	milliard
12	12	douze	30	XXX	et
13	13	treize	31	0	zéro
14	14	quatorze	32	80	quatre-vingts
15	15	quinze	33	100	cents
16	16	seize	34	10^6	millions
17	17	dix-sept	35	10^9	milliards
18	18	dix-huit			

Pour effectuer cette conversion, on utilise, en quelque sorte, une décomposition en base 1000 des nombres ainsi que certaines transformations mineures du nombre lui-même, permettant une adaptation du résultat aux éléments de la liste de mots précédente. En annexe sera détaillé l'algorithme conduisant à cette conversion, qui constitue une partie indépendante du programme de transcription, et dont le rôle est de lui fournir des mots à transcrire.

1.3.5. Liaisons - élisions - dénasalisation

Comme introduction à ce paragraphe, on peut citer un passage du livre de R. Thimonnier (p. 305), qui mettra en évidence l'importance de deux de ces phénomènes :

"Il est vrai que définir les rapports actuels de la prononciation et de l'écriture pose un difficile problème (...). On s'apercevra que la complexité du problème tient surtout à l'existence d'un phénomène assez mal connu - encore que très caractéristique du français moderne - à savoir l'instabilité du mot phonétique.

Celle-ci est due surtout à deux causes, de nature diverse mais étroitement connexes : la liaison et l'élision".

D'autre part, on ne désire pas que la transcription d'un texte revienne à la transcription successive de mots isolés. Or, tenir compte des liaisons, des élisions, de la dénasalisation est une manière de résoudre le problème de la transcription d'un texte. En effet, un mot n'est plus considéré comme isolé, mais comme étant suivi d'un autre mot (si c'est le cas dans le texte en question). Il est inutile de tenir compte du mot précédent, s'il existe, car il est généralement admis que sa présence ne modifie pas la prononciation du mot en question.

Mais pour ce faire, il n'est sans doute pas nécessaire de procéder à plusieurs examens successifs d'un texte que l'on transformerait peu à peu. La solution adoptée, permettant de répondre aux phénomènes de liaison, élision, dénasalisation est simple, et s'insère dans l'unique examen du texte à transcrire. Elle est constituée de deux phases :

- la première est tout à fait inaccessible à l'utilisateur car elle fait partie intégrante du programme. Elle décide de la possibilité ou de l'impossibilité de liaison - élision - dénasalisation (L.E.D.) d'une manière grossière : si un mot est suivi d'un élément de séparation autre qu'un espace ou un trait d'union, aucune L.E.D. n'est effectuée qui dépende du mot suivant. Ceci est confirmé, en ce qui concerne les liaisons, dans l'ouvrage de M. Grévisse, "Le bon usage" : (p. 58, 101, 1er).

"la liaison ne peut avoir lieu (...) dans la lecture, après un signe de ponctuation". De plus, si le dernier phonème du mot précédent est identique à celui que propose la liaison, quelques soient les règles, celle-ci est annulée.

Exemples de liaisons

Petit enfant

P Ø E T I T A N F A N

Huit enfants

Y I T A N F A N

- la seconde étape, est accessible à l'utilisateur. En effet, dans le cas d'une L.E.D. déclarée possible au cours de l'étape précédente, elle consiste en l'application de règles appartenant à une grammaire de transcription précédemment décrite. Il est à noter que les règles d'élision et de dénasalisation sont appliquées lors de la transcription même du mot, et sont sujettes aux mêmes conditions de validation (cf 2-1 ch II) que toute autre règle . Cependant, un cas de liaison possible entre deux mots ne sera étudié qu'après achèvement de la transcription du premier mot de la paire. Une règle sera validée dans les mêmes conditions que ci-dessus, mais si aucune règle concernant les liaisons n'est validée, aucune erreur ne sera signalée, contrairement aux règles concernant la transcription de mots, et on considèrera qu'aucune liaison ne peut se faire dans ce cas.

Cependant certaines réserves sont à émettre, en ce qui concerne les règles relatives aux liaisons :

a) - Elles ne tiennent pas compte de la syntaxe.

A cet effet, on peut citer ici, une phrase de l'ouvrage de Benveniste et Chervel - "l'orthographe" - (p. 132) qui marque bien la dépendance du phénomène de liaison vis à vis de la syntaxe :

"C'est que la liaison opère dans certaines tournures grammaticales seulement".

Or la grammaire de transcription dont on dispose ne tient compte que de l'orthographe, et pour le moment, on ne dispose pas toujours de programmes effectuant l'analyse syntaxique d'un texte. Si ceci était réalisé, il serait alors possible de pallier à cette lacune de la grammaire en ajoutant à toute règle un ou deux éléments supplémentaires, stipulant la catégorie syntaxique des contextes gauche ou droit

exemple :

si P signifie pronom

si V signifie verbe

$S \langle P \rangle + \langle V \rangle$ "vowel" = $[Z _]$

Cette règle pourrait signifier que l'on doit faire la liaison entre tout pronom et tout verbe, si le premier se termine par un S et si le second commence par une voyelle.

b) - la dimension de leurs contextes n'est pas adaptée à ce genre d'application.

Tout d'abord, il faut rappeler ici, que les contextes gauche et droit d'une règle sont limités quant au nombre des éléments qui peuvent y figurer. Or ceci, de même que la présence d'une majorité de règles à contexte gauche inexistant ou très court, a poussé à choisir la possibilité de mettre plus d'éléments dans le contexte droit d'une règle que dans le contexte gauche. Or (cf. 2-2-2-2, , ch. II), les règles de transcription concernant les liaisons appartiennent plutôt à la minorité : c'est-à-dire qu'elles ont un contexte gauche plus important que leur contexte droit qui, sauf exception, se réduit au terme "vowel". Si ceci ne nuit en aucun cas au classement des règles, il est difficile parfois de ne pas trop généraliser l'application de certaines règles :

exemple : $CHEZ + _ +$ "vowel" = $[Z _]$

Supposons que cette règle ait été insérée dans la grammaire pour signifier une liaison obligatoire (cf. John Agraen, p. 5) entre la préposition "chez" et un mot, suivant immédiat, commençant par une voyelle. La longueur imposée aux contextes interdit dans ce cas de préciser que le caractère 'C' est le début du mot ($\# CHEZ$). Aussi, la liaison va être rendue obligatoire entre tout mot se terminant par "CHEZ" et tout mot, suivant immédiat, commençant par une voyelle, si des règles interdisant ces liaisons ne figurent pas dans la grammaire, en position adéquate :

exemple : cachez onze louis d'or
(Z)

dans ce cas la liaison se fera, si la règle suivante (ou tout autre règle équivalente à celle-ci dans ce cas particulier)

"conso" + [] + ONZE # = [[]]

n'est pas examinée avant la règle en question : dans ce cas très particulier, l'insertion d'une règle incomplète, rendant une liaison obligatoire, a conduit à effectuer une liaison interdite (cf. John Agree, p.6). Mais ceci n'est tout de même pas fréquent, et dans la majorité des cas on peut parer à de tels incidents.

1.3.6. Erreurs

Au cours de la transcription d'un texte, il peut se produire des erreurs. Ceci n'arrête en aucun cas la transcription du texte. Seulement, si un mot quelconque ne peut être entièrement transcrit, la transcription partielle du mot en question ne sera pas imprimée dans les résultats, mais l'erreur sera signalée par des astérisques (** **).

I L R Ø E G A R D Ø E L Ø E ** ** T R A V A I E E .

1.3.7. Présentation des résultats

L'intelligibilité du texte transcrit, voilà ce qui a été à l'origine de toutes les dispositions prises.

a) Généralités

La transcription d'un texte ne se fait pas après une phase de décomposition syllabique de ce même texte (la syllabation est différente selon que le mot est isolé, ou réalisé dans la chaîne parlée (en contexte) ex : cher - C H A I R - ; cher ami - C H A I - R A - M I . Aussi, dans les résultats apparaîtront isolés tous les mots du texte (mot, dans le sens défini en 1.3.2., ch. III) ainsi que les mots provenant de la conversion nombre - graphèmes, et les liaisons (cf. Gsell S. "16 transcriptions phonétiques de textes de prose et de vers")

b) Distinction aisée entre les mots et les phonèmes d'un même mot

Ceci est résolu par un espace de séparation nettement supérieur entre les mots (4 espaces) qu'entre les phonèmes (1 espace). De plus, les phonèmes, qui dans l'écriture des règles comptent tous deux caractères, ne comportent plus de caractère 'L', pour éviter toute confusion entre les espacements. Ainsi, la séparation, entre mots et entre phonèmes, est régulière tout au long du texte transcrit, et facilite la lecture des résultats.

c) Ponctuation

Afin de mieux se repérer dans un texte transcrit, l'utilisateur peut demander que la ponctuation du texte à transcrire figure dans le texte transcrit. Tous les signes de ponctuation peuvent y être insérés au fur et à mesure de la transcription, sauf le trait d'union. En effet, celui-ci n'étant pas un obstacle pour la liaison, la présence consécutive du trait d'union et d'une liaison diminue quelque peu la signification de cette dernière.

d) Aucun mot transcrit n'est coupé : il se trouve entièrement sur la même ligne. Ceci donne lieu à des lignes de longueur variable. Une lecture plus facile du résultat entraîne donc un léger défaut dans la présentation de ce dernier, mais en augmente l'intelligibilité.

Il faut noter que le temps d'exécution de toute transcription est imprimé en fin du texte. Ce temps est compté sur une horloge dont la précision est de 2/100 de seconde. Le temps d'accès au fichier contenant le texte à transcrire en est exclu, d'une manière correspondant à la précision de l'horloge en question. De plus, plusieurs textes contenus dans les fichiers différents, peuvent être transcrits successivement (il suffit de répondre à une question posée à la console) sans pour autant relancer le programme présentement utilisé.

K I W A T O N K " I L F O E R A B A U
D O E H I N , A P E A I D O E H I N ?

Extrait de :

"LA MARE AU DIABLE".G.SAND

Germain, lui dit un jour son beau-père, il faut pourtant te décider à reprendre femme. Voilà bientôt deux ans que tu es veuf de ma fille, et ton aîné a sept ans. Tu approches de la trentaine, mon garçon, et tu sais que, passé cet âge-là, dans nos pays, un homme est réputé trop vieux pour entrer en ménage. Tu as trois beaux enfants, et jusqu'ici ils ne nous ont point embarrassés. Ma femme et ma bru les ont soignés de leur mieux et les ont aimés comme elles le devaient.

J A I R M I N , L Y I . D I T U N J O U R S O N
B A U P A I R , I L F A U P O U R T A N T O E
D E S I D E A R O E P R A N D R O E F A M . V W A L A
B I E I N T A U D E U Z A N K O E T U A I V O E F
D O E M A F I I E , E T O N N A I N E A
S A I T A N . T U A P R O C H D O E L A
T R A N T A I N , M O N G A R S O N , E T U S A I
K O E , P A S E S A I T A J L A , D A N N A U
P A I I , U N N O M A I R E P U T E T R O
V I E E U P O U R A N T R E A N M E N A J . T U A
T R W A B A U Z A N F A N , E J U S K " I S I
I L N O E N O U Z O N P W I N T A N B A R A S E
. M A F A M E M A B R U L A I Z O N
S W A G N E D O E L O E R M I E E U , E L A I Z
O N T A I M E K O M A I L L O E D O E V A I .

CONCLUSION

Contrairement à la construction d'une grammaire, la transcription d'un texte, qui est en fait l'exploitation d'un effort fourni au préalable, ne doit pas demander de travail à l'utilisateur, mais refléter exactement l'emploi d'un outil qu'il a mis au point (une grammaire). C'est ce qui a été tenté, afin de pallier à une indépendance complète de ces deux parties, due à la fois à la taille - mémoire du mini - ordinateur et au peu d'intérêt porté au gain en place - mémoire. En effet, une indépendance complète, quant à la forme, pour deux programmes (ce qui est le cas), accompagnée d'une dépendance marquée quant à leur utilisation (ce qui n'est pas le cas), obligerait l'utilisateur à passer incessamment de l'un à l'autre, et donc à perdre du temps. Mais, la conception de chacune des deux parties présentées permet de l'éviter, et donc leur indépendance ne nuit pas à leur efficacité.

2- ETUDE STATISTIQUE ELEMENTAIRE DES REGLES DE TRANSCRIPTION.

Après transcription d'un ou plusieurs textes, il est possible, mais non imposé, d'obtenir l'impression de résultats chiffrés, concernant l'emploi des différentes règles de la grammaire de transcription.

2-1) Présentation.

Pour chacune des règles de la grammaire de transcription, employée dans ce cas particulier, sont donnés :

- l'indice de la règle (colonne A)
- le nombre de fois qu'elle a été employée dans la transcription (colonne C)
- son pourcentage d'utilisation dans la transcription (colonne B)

Ces trois éléments apparaissent en colonnes dans un tableau dont chaque ligne correspond à une règle. (exemple page suivante).

Si une règle de la grammaire n'apparaît pas dans ce tableau, cela signifie qu'elle n'a jamais été utilisée au cours de la transcription précédente. De plus, il faut noter que le rangement des règles dans ce tableau, correspond à l'ordre décroissant des éléments des colonnes B et C.

A la suite de ce tableau sont donnés deux nombres qui peuvent être utilement consultés avant d'évaluer les statistiques élémentaires effectuées ; ce sont :

- le nombre total de règles employées,
- le nombre total de règles différentes employées.

Col. A

Col. B

Col. C

INDICE DE LA REGLE	% UTILISATION	NBRE D EMPLOIS
I 130	I 5.70	I 17 I
I 147	I 4.70	I 14 I
I 141	I 4.36	I 13 I
I 21	I 4.03	I 12 I
I 100	I 3.36	I 10 I
I 214	I 3.36	I 10 I
I 68	I 3.02	I 9 I
I 96	I 2.68	I 8 I
I 104	I 2.68	I 8 I
I 123	I 2.68	I 8 I
I 145	I 2.68	I 8 I
I 153	I 2.35	I 7 I
.
I 173	I 0.34	I 1 I
I 206	I 0.34	I 1 I
I 216	I 0.34	I 1 I
I 228	I 0.34	I 1 I
I 231	I 0.34	I 1 I
I 235	I 0.34	I 1 I
I 236	I 0.34	I 1 I
I 237	I 0.34	I 1 I
I 238	I 0.34	I 1 I
I 242	I 0.34	I 1 I
SUR UN TOTAL DE DONT	298 92	REGLES UTILISEES REGLES DIFFERENTES

remarque : ces comptages sont limités par le mini ordinateur employé. Ainsi aucun de ces nombres entiers ne pourra excéder $2^{15} - 1$, soit (32767).

2-2) Utilisation.

Les résultats chiffrés donnés à la suite d'une transcription sont très simples et se réduisent à des comptages ou des pourcentages. Si aucun résultat un peu plus particulier n'a été donné, c'est pour ne pas en diminuer le champ d'application, et parce que, dans la première partie (cf. ch. III), il est possible d'obtenir des renseignements aidant à mener un calcul plus précis.

Quelles correspondances existe-t-il entre ces deux étapes-résultats, à savoir les statistiques élémentaires sur l'emploi des règles d'une grammaire de transcription et les renseignements que l'on peut obtenir sur cette grammaire ?

a) Dans le tableau relatif aux statistiques élémentaires, toute règle est reprise par un numéro : son indice. Or, il est possible d'obtenir la liste, classée suivant les indices croissants, de toutes les règles de transcription d'une grammaire (ordre 15 du tableau "renseignements"). Ceci permet de faire la relation entre l'indice et le texte d'une règle. A partir du texte de la règle, il est possible de situer celle-ci dans le classement des règles, relatif au fonctionnement de la grammaire (§ 2-2, ch. III), et par là retrouver son indice, si c'est nécessaire. Mais il faut noter que le texte d'une règle est toujours accompagné de de l'indice de celle-ci. L'inverse n'est pas exact.

b) D'autre part, si l'on veut déterminer le nombre de fois que l'on a utilisé des règles appartenant à une certaine catégorie, il n'est pas toujours nécessaire de rechercher dans la grammaire entière les indices des règles qui intéressent l'utilisateur. En effet, il est possible d'obtenir automatiquement les textes et les indices des règles:

- comptant M caractères dans le terme médian (ordre 7)
- comptant M termes dans le contexte gauche ou droit

(ordre 10 et 11)

- donnant un phonème particulier (ordre 13).

Statistiques effectuées sur l'extrait de "La Mare au Diable"
présenté p. 60.

INDICE DE LA REGLE		% UTILISATION		NOMBRE D'EMPLOIS	
I	7	I	14.21	I	54
I	29	I	8.42	I	32
I	22	I	7.89	I	30
I	28	I	7.63	I	29
I	1	I	7.11	I	27
I	30	I	6.58	I	25
I	27	I	5.53	I	21
I	15	I	5.26	I	20
I	23	I	5.26	I	20
I	21	I	4.47	I	17
I	37	I	3.95	I	15
I	20	I	3.68	I	14
I	25	I	3.16	I	12
I	6	I	2.63	I	10
I	8	I	2.37	I	9
I	12	I	1.58	I	6
I	4	I	1.32	I	5
I	5	I	1.32	I	5
I	13	I	1.32	I	5
I	32	I	1.05	I	4
I	34	I	1.05	I	4
I	2	I	0.79	I	3
I	26	I	0.79	I	3
I	14	I	0.53	I	2
I	18	I	0.53	I	2
I	3	I	0.26	I	1
I	9	I	0.26	I	1
I	16	I	0.26	I	1
I	24	I	0.26	I	1
I	35	I	0.26	I	1
I	38	I	0.26	I	1
SUR UN TOTAL DE		380	REGLES UTILISEES		
DONT		31	REGLES DIFFERENTES		

Grammaire utilisée pour ces statistiques

		REGLE D INDICE: 12	
		+F+=[]	
REGLE D INDICE: 2			
+A2+=[]		REGLE D INDICE: 13	
REGLE D INDICE: 3		+G+=[]	
+A3+=[]			
REGLE D INDICE: 1			
+A+=[]		REGLE D INDICE: 14	
		+H+=[]	
REGLE D INDICE: 4			
+B+=[]		REGLE D INDICE: 16	
		+I3+=[]	
REGLE D INDICE: 38		REGLE D INDICE: 17	
+C5+=[]		+I4+=[]	
REGLE D INDICE: 5		REGLE D INDICE: 15	
+C+=[]		+I+=[]	
REGLE D INDICE: 6		REGLE D INDICE: 18	
+D+=[]		+J+=[]	
REGLE D INDICE: 8		REGLE D INDICE: 19	
+E1+=[]		+K+=[]	
REGLE D INDICE: 9			
+E2+=[]		REGLE D INDICE: 20	
REGLE D INDICE: 10		+L+=[]	
+E3+=[]			
REGLE D INDICE: 11			
+E4+=[]		REGLE D INDICE: 21	
REGLE D INDICE: 7		+M+=[]	
+E+=[]			

REGLE D INDICE: 22
+N+=[]

REGLE D INDICE: 24
+O3+=[]

REGLE D INDICE: 23
+O+=[]

REGLE D INDICE: 25
+P+=[]

REGLE D INDICE: 26
+Q+=[]

REGLE D INDICE: 27
+R+=[]

REGLE D INDICE: 28
+S+=[]

REGLE D INDICE: 29
+T+=[]

REGLE D INDICE: 31
+U2+=[]

REGLE D INDICE: 30
+U+=[]

REGLE D INDICE: 32
+V+=[]

REGLE D INDICE: 33
+W+=[]

REGLE D INDICE: 34
+X+=[]

REGLE D INDICE: 35
+Y+=[]

REGLE D INDICE: 36
+Z+=[]

REGLE D INDICE: 37
+ +=[]

Dans les cas cités ci-dessus, ou encore dans d'autres éventualités, il ne restera à l'utilisateur qu'à effectuer des additions, afin d'obtenir les statistiques élémentaires relatives à un ensemble de règles.

Comme exemple d'utilisation de ces statistiques va être présenté le comptage des caractères alphabétiques d'un texte.

2-3) Application particulière :

Recherche sur l'orthographe du français.

On peut mettre au point plusieurs types de règles : certaines règles plus "empiriques" que d'autres, auront pour but de transcrire de manière la plus "économique" possible, les suites graphiques en suites de phonèmes :

- ex: 1) + O S + "vowel" = [AUZ_U] (rose)
2) + EILL + = [AIIE] (réveille)
3) + IEN + "CNONN" = [IEIN] (bientôt)

Ces règles sont qualifiées "d'économique", car elles transcrivent un nombre maximal de caractères orthographiques, et donc diminuent le nombre de règles à examiner dans ces cas particuliers, contrairement à l'exemple qui va suivre.

En effet, il est possible de déterminer des règles plus "fines" mais plus générales qui donnent des renseignements sur la nature et le fonctionnement de l'orthographe, tout en permettant une aussi bonne phonétisation.

Dans une grammaire, les trois règles de transcription précédentes pourraient être remplacées par les six règles suivantes (deux règles remplaçant une règle "économique") :

- 1- + O + (S + "vowel") = [AU]
"vowel" + S + "vowel" = [Z_U]

$$\begin{array}{lll} 2- & + EI + LL & = [AI] \\ & i + LL + & = [IE] \\ \\ 3- & + i + "vowel" & = [IE] \\ & I + EN + "CNONN" & = [IN] \end{array}$$

remarque : une règle porte sa réelle signification, seulement si elle apparaît dans le sous-ensemble classé de la partition des règles, (ou liste) auquel elle appartient. Ceci n'est pas le cas ci-dessus.

Cette liberté dans le choix des règles de grammaire intéresse en particulier Vincent LUCCI, (Maître Assistant à l'Institut de Phonétique de Grenoble). Voici un résumé de ses préoccupations à propos de l'orthographe du français :

Les descripteurs de l'orthographe du français ont l'habitude de dire que cet orthographe est alphabétique, ou encore qu'une composante de l'orthographe est phono-graphique (c'est-à-dire que les lettres renvoient à la prononciation) et qu'une autre partie renvoie directement au sens par l'intermédiaire de lettres non prononcées :

ex : pied, parlent

Les règles utilisées dans un programme de phonétisation permettent d'apporter un complément statistique à la description de chacune de ces composantes.

La composante phono-graphique:

a) les rapports lettres-son peuvent être directs :

ex : la lettre " J " renvoi au phonème 3, ou "Y" a V.

Mais la plupart du temps, les correspondances sont indirectes, c'est-à-dire qu'une même lettre (ex: c, g, t,) ne renvoie pas à la même prononciation selon les lettres qui suivent, ou bien des suites de deux lettres

(an, ll, etc...) renvoient à des phonèmes différents selon l'entourage graphique (ex: ban, vane ; molle, fille).

On peut étudier ces "distorsions" (terme de GAK) et voir ce que chacune d'elle représente en pourcentage. L'étude statistique apportera un complément quantitatif aux études fondamentales de cette composante phono-graphique, et donc à celle de l'orthographe du français en général.

b) Les règles ont été élaborées de manière à prendre en compte les principales valeurs phono-graphiques (ou types de correspondances lettres-phonèmes, ou encore les types de "distorsions") dont parlent les spécialistes de l'orthographe, et qui sont :

<p>Monographe de base</p>	<p>a) une lettre renvoie à un phonème (abstraction faite de l'entourage graphique) $+ A + = [A_{\square}]$</p> <p>b) une lettre renvoie à deux phonèmes (abstraction faite de l'entourage graphique) : $+ X + = [K_{\square} S_{\square}]$</p>
<p>Monographe de position</p>	<p>a) Une lettre renvoie à un phonème dans un contexte particulier $+ C + "E - I - Y" = [S_{\square}]$ $+ O + IN = [W_{\square}]$</p> <p>b) Une lettre renvoie à deux phonèmes dans un contexte particulier $E + X + "vowel" = [G_{\square} Z_{\square}]$ $+ O + (Y + "vowel") = [W_{\square} A_{\square}]$</p>

<p>Digramme de base</p>	<p>a) Deux lettres renvoient à un phonème (abstraction faite de l'entourage graphique) :</p> <p>+ AI + = [AI] + GN + = [GN]</p> <p>b) Deux lettres renvoient à deux phonèmes (abstraction faite de l'entourage graphique) :</p> <p>+ OI + = [wɔi]</p>
<p>Digramme de position</p>	<p>a) Deux lettres renvoient à un phonème dans un contexte particulier</p> <p>+ CH + "conso" = [k] + EZ + # = [z]</p> <p>b) Deux lettres renvoient à deux phonèmes dans un contexte particulier</p> <p>+ AY + "vowel" = [aij] + AY + S = [aj]</p>
<p>Trigramme de base</p>	<p>Trois lettres renvoient à un phonème (abstraction faite de l'entourage graphique) :</p> <p>+ EAU + = [o] + OEU = [œ]</p>
<p>Trigramme de position</p>	<p>Trois lettres renvoient à un phonème dans un contexte particulier.</p> <p>A + ILL + "vowel" = [ij] + AIN + "CNONN" = [in] + SCH + = [ʃ]</p>

c) Lorsqu'une lettre ne correspond à rien dans la prononciation (valeur "zéro"), il est utile de savoir s'il s'agit :

- d'une consonne finale (chaud)
- de deux consonnes finales (poids)
- d'une voyelle finale (pie)
- d'une consonne non finale (doigté)
- d'une voyelle non finale (engouement, geai)

Ces lettres non prononcées fournissent des renseignements en général sur la composante "idéographique" ou "sémiotique".

APPLICATIONS :

a) En pédagogie.

Il faut déterminer dans quelle mesure exacte l'orthographe du français est phono-graphique, et comment, pour pouvoir mieux l'enseigner. A l'heure actuelle, certains enseignants (J.Foucambert, E. Charmeux) disent que l'orthographe du français est essentiellement "idéographique" et que, dans l'enseignement, il faut apprendre globalement (comme un idéographe) à écrire le mot, sans passer par un apprentissage des correspondances lettres-son. Or, ceci n'est pas confirmé statistiquement.

- En pédagogie, un principe important : apprendre à l'élève, en priorité ce qui est le plus fréquent, car le plus utile. Là encore, on pourrait souhaiter une progression dans une méthode de lecture ou d'écriture, fondée sur la fréquence des correspondances phono-graphiques.

b) Pour la lecture.

- Certaines conceptions pédagogiques de la lecture prônent une non mise en correspondance des lettres et des phonèmes.

Dans la lecture, l'oeil perçoit, par des balayages très rapides,

le sens d'un texte à partir d'indices sélectionnés (lettres non prononcées). Ces indices peuvent être de type lexical (ex: doigt), de type morphologique (ex: grand) ou de type grammatical (ex: parlaient). Il serait utile de savoir quelles proportions représentent ces indices (lettres non prononcées): indices morphologiques, lexicaux, grammaticaux.

- Les règles du programme de phonétisation permettent d'apporter des précisions quantitatives non négligeables. On peut ainsi évaluer la part de " redondance " à l'écrit.

Ces recherches peuvent servir par ailleurs à l'enseignement (à la mode) de la lecture rapide, et peut être même de la sténographie.

ETUDE - CONCLUSION

Diversification des résultats

L'étude présentée dans les deux chapitres précédents concerne uniquement la transcription de textes, c'est-à-dire le passage d'une chaîne de caractères orthographiques à une chaîne de caractères phonétiques (codés dans le cas présent).

1°) Syllabation

Or, la transcription d'un texte peut se faire suivant deux méthodes :

- soit, elle est précédée d'une décomposition syllabique des mots du texte (cf. N. Catach et V. Messonnier, Galf 79, p. 178-179-180-181-182)
- soit, elle se fait directement, à partir du texte orthographié (par ex. cf. Fervers, Leroux, Miclet).

Il serait donc intéressant de pouvoir réaliser la décomposition syllabique des mots d'un texte et transcrire le texte résultat lors d'un second passage en machine à l'aide de grammaires de syllabation et de transcription, qui, dans l'un et l'autre cas, seraient construites tout comme les grammaires déjà mentionnées. Pour ce faire, seuls interviendront quelques changements à propos des caractères autorisés ou interdits dans les règles, ainsi que dans la présentation des résultats.

a) Grammaire

Le texte résultant d'une syllabation comporte tous les caractères existant dans les mots du texte origine, auxquels est ajouté un symbole séparateur de syllabes. Or on sait que ne pourront figurer dans les résultats que le ou les caractères formant les phonèmes présentés p. 11, ch. II. Comment faire alors, pour produire une syllabation correcte du mot "coquille", qui serait : CO - QUI - LLE, si les lettres du mot C et Q ne figurent

dans aucun de ceux-ci ? Pour cette raison, certains éléments, qui ne sont pas des phonèmes ont été ajoutés à la liste de ces derniers, afin que toutes les lettres alphabétiques et les symboles diacritiques puissent être représentés dans le résultat, de même qu'un symbole séparateur de syllabes ("-").

Liste de ces éléments :

lettres alphabétiques	représentation
Q	[Q _u]
X	[X _u]
H	[H _u]
C	[C _u]
symboles diacritiques	représentation
1 (accent aigu)	[1 _u]
2 (accent grave)	[2 _u]
3 (accent circonflexe)	[3 _u]
4 (tréma)	[4 _u]
5 (cédille)	[5 _u]
symboles séparateur de syllabes :	représentation
-	[- _u]

b) Modifications dans la présentation des résultats :

Le résultat représente un texte dont les mots, formés de caractères alphabétiques et encore de symboles diacritiques, sont découpés en syllabes. Ceci est la seule différence du résultat avec le texte de départ. On ne peut donc pas, au risque de confusions, espacer les caractères,

d'un même mot , comme cela est fait par les phonèmes, dans le cas d'une transcription (§ 1.3.5.a., ch. III).

D'autre part, le texte syllabisé devra être présenté de telle façon qu'il puisse être lui-même transcrit lors d'un deuxième passage en machine : ceci impose certaines restrictions qui seront étudiées ci-après (cf. § 2.1.3., ch. V).

GER-MAIN , LU-I DIT UN JOUR SØN BEAU PE2-RE , IL
FAUT PØUR-TANT TE DEI-CI-DER A2 RE-PREN-DRÈ FE-MME .
VØI-LA2 BIEN-TØ3T DEUX ANS QUE TU ES VEUF DE
MA FI-LLE , ET TØN AI3-NE1 A SEPT ANS . TU
A-PP-RØ-CHES DE LA TREN-TAI-NE , MØN GAR-C5ØN , ET
TU SAIS QUE , PA-SSE1 CET A3-GE LA2 , DANS NØS
PAYS , UN HØ-MME EST RE1-PU-TE1 TRØP VIEUX PØUR
EN-TRER EN ME1-NA-GE . TU AS TRØIS BEAUX EN-FANTS }
ET JUS-QU " I-CI ILS NE NØUS ØNT PØINT
EM-BA-RRA-SSEIS . MA FE-MME ET MA BRU LES ØNT
SØI-GNEIS DE LEUR MIEUX , ET LES ØNT AI-MEIS
CØ-MME E-LLES LE DE-VAIENT .

TEMPS DE TRANSCRIPTION : OMN 8" 72/100

c) Transcription du texte syllabisé

Celle-ci se fera à l'aide d'une grammaire de transcription ordinaire. L'élément séparateur de syllabe (-), autorisé dans les contextes gauche et droit d'une règle appartenant à une grammaire quelconque (afin d'éviter de poser trop de questions à l'utilisateur), sera utile dans ce cas :

	<u>mot</u>	<u>mot syllabisé</u>	<u>mot syllabisé transcrit</u>
ex :	RUSE1 (rusé)	RU - SE1	R U Z E

2°) Transcription syllabique

D'autre part, comme on l'a noté précédemment (cf. § 1.3.5a), la syllabation d'un texte varie, selon qu'un mot du texte est ou non considéré comme mot isolé. Etant donné, que dans l'éventualité de syllabation, présentée dans le paragraphe précédent, chaque mot d'un texte est considéré isolément, il serait bon de disposer de moyens permettant "une transcription syllabique" d'un texte. Tout d'abord, qu'est-ce qu'une transcription syllabique ? : c'est une transcription phonétique qui tient compte des frontières syllabiques dans le texte, et non plus dans un mot du texte :

- J A I R - M A I N , - L U - I - D I - T U N - J Ø U R - S Ø N -
- B A U - P A I R - , I L - F A U - P Ø U R - T A N - T Ø E -
- D E - S I - D E - A - R Ø E - P R A N - D R Ø E - F A M - .
- V U A - L A - B I E I N - T A U - D E U - Z A N - K Ø E - T U -
- A I - V Ø E F - D Ø E - M A - F I I E - , E - T Ø N - N
A I - N E - A - S A I - T A N - . - T U - A - P R Ø C H - D Ø E -
- L A - T R A N - T A I - N , - M Ø N - G A R - S Ø N - , E -
- T U - S A I - K Ø E , - P A - S E - S A I - T A - J - L A -
, - D A N - N A U - P A I - I - , U N - N Ø M - A I
- R E - P U - T E - T R Ø P - V I E U - P Ø U - R A N - T R E - A N
- M E - N A J - . - T U - A - T R W A - B A U - Z A N - F A N -
, E - J U S - K I - S I - I L - N Ø E - N Ø U - Z Ø N -
- P W - I N - T A U - B A - R A - S E - . - M A - F A M - E -
- M A - B R U - L A I - Z Ø N - S W A - G N E - D Ø E - L E U R
- M I E U - , E - L A I - Z Ø N - T A I - M E - K Ø M - A I - L -
L Ø E - D Ø E - V A I .

a) Grammaire

La grammaire permettant une transcription syllabique, demande peu de chose en plus de ce qui est imparti à une grammaire de transcription ordinaire. En effet, seul sera ajouté à la liste des phonèmes présentée auparavant (cf. § 1.2.3 ch. II), l'élément séparateur de syllabes : "-". Mais ceci ne signifie pas pour autant que la mise au point de l'une ou l'autre de ces grammaires demande le même travail. Si dans une grammaire de transcri-

ption ordinaire, élisions, liaisons et phonétisation sont indépendantes, il n'en est plus de même pour une grammaire de transcription syllabique, qui en plus marquera les frontières de syllabes. On peut prendre en exemple un cas d'élision du 'e' :

<u>transcription syllabique</u>	
petite	a) aucune élision - P OE - TI - TOE -
	b) élision du 'e' final - POE - TIT -
	c) élision du premier 'e' - PTI - TOE -
	d) élision des deux lettres 'e' - PTIT -

Sur cet exemple on voit que le nombre de syllabes du mot 'petite' varie avec les élisions de la voyelle 'e'. Il n'y a donc plus indépendance de certaines règles, ce qui rend plus délicate la construction d'une grammaire de transcription syllabique, mais non impossible.

b) Présentation du résultat

Les seules différences entre un texte transcrit ordinairement, et un texte transcrit "syllabiquement" sont les suivantes :

- marquage des frontières syllabiques
- pas d'espacement entre les mots non séparés par un signe de ponctuation (autre que l'apostrophe ou le trait d'union).

3°) Transcription libre

Par ce terme, on désigne en fait toute "transcription" pour laquelle on va utiliser une grammaire moins "spécialisée" que les autres : en effet, on disposera pour sa construction de tous les phonèmes présents §1.2.3 ch. II, ainsi que de tous les éléments que l'on a ajouté à cette liste (cf. § 1 - ch. III). De plus, comme pour la décomposition syllabique, il ne sera pas laissé d'espaces entre les éléments du résultat, si celui-ci est, à la demande de l'utilisateur, inscrit en mémoire afin de permettre un autre passage en machine.

REALISATION INFORMATIQUE

Introduction

La réalisation informatique de ce projet dépend principalement :

1°- du matériel utilisé et de son mode d'utilisation :

Il s'agit ici d'un miniordinateur DIGITAL, le LSI 11/2, qui ne peut être utilisé que par une seule personne à la fois. Il s'ensuit que la place mémoire disponible pour un programme donné est égale à la place mémoire totale (24 K octets). Dans un programme il ne sera donc pas nécessaire d'économiser cette place mémoire pour en laisser une partie innocupée.

2° - de la finalité de ce projet :

Celui-ci n'étant pas conçu pour une quelconque application industrielle, mais pour devenir plutôt un instrument de recherche, la compétitivité de son coût, c'est-à-dire principalement, du temps d'exécution des programmes, n'est pas primordiale.

D'autre part, cela tient aussi au fait qu'un seul utilisateur peut se servir du mini-ordinateur : pas de travail en temps partagé, pas d'emcombement du système qui soit gênant pour d'autres éventuels utilisateurs. De plus, l'impression des résultats se fait sur une imprimante lente (10 caractères par seconde), ce qui fausse quelque peu la notion de vitesse d'exécution d'un programme. Enfin, l'horloge comptant un temps d'exécution, a une précision de $2/100$ de seconde, on ne cherchera donc pas à grignoter quelques micro-secondes, ni même quelques centièmes de seconde.

3°- des principaux souhaits de l'utilisateur, à savoir :

- commodités d'utilisation - clarté du programme ; fonctionnement et résultats.

- adaptation à ceux qui ne sont pas des familiers de l'informatique.
- pas de compilation d'une grammaire de transcription après modification.

- diversification des résultats : transcription
 - comptage de phonèmes ou graphèmes
 - décomposition syllabique des mots d'un texte.
 - transcription syllabique d'un texte.

* Finalement, cette réalisation informatique ne sera pas une recherche de vitesse d'exécution des programmes, ni d'économie de la place mémoire utilisée, mais, tenant compte de son cadre d'utilisation, elle sera plutôt une réponse aux souhaits d'éventuels utilisateurs.

CHAPITRE IV

- LA GRAMMAIRE

Les raisons principales, à l'origine de sa structure informatique, sont les suivantes :

- on veut qu'une grammaire soit utilisable, après toute modification, sans compilation préalable.(interpréteur).
- on veut que ces modifications soient aisées.

Ceci a entraîné la représentation de la grammaire par un tableau d'entiers, dont la taille a été imposée par la place mémoire disponible. Chaque règle de la grammaire occupe une ligne de ce tableau (N lignes, M colonnes), et peut être repérée par la seule indication du numéro de cette dernière, que l'on appelle ainsi "indice de la règle".

-] Dimensions de la grammaire

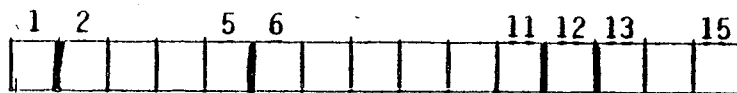
Ces dimensions sont en fait les dimensions du tableau qui la contient : $I_{gram}(N,M)$.

N représente le nombre de lignes du tableau, ou encore le nombre

maximum de règles que peut contenir une grammaire.

M représente le nombre d'éléments que peut, au maximum, comporter une règle quelconque, auxquels sont ajoutés deux éléments intervenant dans le fonctionnement de la grammaire. Mais afin de reconnaître le "squelette" d'une règle, (cf. p.15 CH II), l'emplacement dans une ligne de certains éléments de la règle sera imposé, ce qui déterminera les longueurs maximales LG et LD (cf. p. 34, ch. II).

Voici le schéma d'une ligne de Igram :



* M = 15

* LG max = 4

* LD max = 6

* emplacements imposés :

1 : indice de la règle suivante dans le classement (cf. p. 32, ch. II)

2..5 : éléments du contexte gauche, convertis en entiers (p. 88, ch. IV)

6..11: caractères transcrits, convertis en entiers, et (p.88)

éléments du contexte droit, convertis en entiers (p.88)

12 : nombre de caractères transcrits

13..15 : phonèmes, convertis en entiers (p.88)

N.B. Dans la suite de de paragraphe, caractères transcrits et contexte droit seront englobés dans le terme contexte droit (CT + CD = CD)

Justification des valeurs maximales de LG et LD.

Dans la grammaire vont figurer, sous la même forme, des règles concernant la transcription de mots ordinaires, d'exceptions et de liaisons. Or ces trois catégories présentent une certaine dissymétrie dans la longueur, en nombre d'éléments, des contextes gauche et droit.
ex:

catégorie	règle	LG	LD	exemple d'applicati
mots ordinaires	+ ou + = [OU]	0	2	bouton
exceptions	#p + ay + s = [AI I _U]	2	3	paysan
liaison	#et + □ + "vowel" = [□□]	3	2	riche et avare

On s'aperçoit dans cet exemple, des variations de LG et LD relatives à ces différents cas :

- pour un mot ordinaire : le contexte gauche est vide, en général.
- pour une exception : les deux contextes ont en général leur importance (il s'agit de distinguer une suite de caractères, donc d'en reconnaître le plus possible).
- pour une liaison: sauf cas particulier, le contexte gauche est plus important que le contexte droit. En effet, dans le cas général, pour qu'une liaison puisse s'effectuer entre deux mots, il faut que le premier réponde à des conditions particulières, mais il suffit que le second commence par une voyelle. Or, ces conditions particulières sont bien souvent des conditions de syntaxe, que l'on va tenter d'assimiler à des conditions orthographiques, dans les cas où cela est possible (pour l'exemple ci-dessous, cf. J. Agreen, p. 5 et 6) :

Liaison obligatoire :

préposition monosyllabique + mot quelconque commençant par
une voyelle

exemple :

en_été, en_avril, en_occident, en_image, en_usine, ...

règle :

$$\#en + \square + \text{"vowel"} = [N_\square]$$

Afin d'unifier ces trois catégories, il est nécessaire de choisir un compromis, en ce qui concerne les dimensions maximales autorisées pour les contextes gauche et droit, compromis qui satisfasse cependant à la majorité des cas.

Il a donc été décidé de tenir compte davantage des règles concernant la transcription de mots ordinaires, que de celles concernant les liaisons, catégories représentant en général les cas extrêmes de dimension du contexte gauche, et du contexte droit :

LG max = 4

LD max = 6

rq: si LG max et LD max sont supérieurs à LG ou LD d'une règle, le tableau est complété par des valeurs 0.

Sur le tableau suivant sera notée la convenance (=), l'insuffisance (-) ou l'excès (+) de ces valeurs, pour chacune des trois catégories de règles :

	mots ordinaires	exceptions	liaisons
LG	+	=	-
LD	=	=	+

Il faut noter que ce tableau a une origine empirique et pas du tout théorique.

De plus, d'autres raisons que celles mentionnées ci-dessus, sont aussi à l'origine du choix de ces valeurs : en effet, la place mémoire, disponible dans l'ordinateur pour l'implantation d'une telle grammaire, peut contenir environ 5000 entiers. Tout en fixant M, il faut aussi fixer N, c'est-à-dire le nombre de règles que peut contenir une grammaire. Celle-ci devant fournir une bonne transcription, le nombre de règles qu'elle peut contenir doit être assez élevé : $N = 320$. Dans la thèse de Divay & Guyomard, (p. 94), le nombre de règles concernant la transcription de mots isolés et de liaisons est de $300 + (30 + 18)$ soit de 348 c'est-à-dire à peu près du même ordre que dans ce travail, mais en fait, il représente un nombre bien supérieur d'éventualités, en raison de la présence des "classes" (cf. Divay & Guyomard, pp 110-111-112). Cependant 320 règles semblent suffire largement au vu des résultats obtenus, et ne conduire qu'à des erreurs plutôt "exceptionnelles", c'est-à-dire peu fréquentes. Ceci est confirmé par la recherche effectuée sur ce sujet par N. Catach, dont certaines conclusions figurent dans l'ouvrage de LUCI et Naze (cf. Lucci et Naze p. 76). En effet 33 groupes de caractères orthographiques seraient seuls utilisés dans 80 à 90 % de la communication orale : (A, E, I, O, U, EU, OU, AN, IN, ON, UN, ILL, Y, OI, OIN, P, B, T, D, C, G, F, V, S, Z, X, CH, J, L, R, M, N, GN). Donc, un nombre de règles voisin suffirait pour une prononciation correcte à 80 % (au maximum) du français.

On s'aperçoit alors que 320×15 représente à peu près la taille mémoire disponible pour l'utilisation d'une grammaire de transcription. Ce sont les dimensions définitives choisies pour le tableau Igram : Igram (320,15).

Pour achever la description d'une règle contenue dans Igram (cf. p. 83) il faut encore mentionner 5 entiers :

1													12	13	14	15
	X	X	X	X	X	X	X	X	X	X	X	X				

- L'entier contenu dans la case de la colonne 1 spécifie l'indice dans Igram de la règle qui lui fait suite dans le classement. Si cet entier est nul, cela signifie que cette règle est la dernière d'une des 27 listes de règles (cf. § 2.2 ch. II).

- L'entier contenu dans la case de la colonne 12, indique le nombre de caractères transcrits par la règle, c'est-à-dire le nombre de caractères dont doit avancer le pointeur correspondant à la chaîne à transcrire, après application de cette règle.

Il est nécessaire d'être en possession de cette information, car dans Igram, CT et CD, c'est-à-dire caractères transcrits et contexte droit ne sont pas séparés, afin de permettre une plus grande facilité de classement et une plus grande liberté de longueur pour CT et CD.

- Enfin, les trois entiers correspondant aux colonnes 13, 14 et 15 représentent les phonèmes résultant de l'application de cette règle : le nombre de 3 peut sembler, dans la plupart des cas, trop important, mais il est parfois nécessaire quand par exemple on veut procéder à une décomposition syllabique des mots : (ce dont on parlera ultérieurement).

ex : $+ CH + = \left[\begin{array}{c} - \\ \cup C \cup H \cup \end{array} \right]$

- Cette règle indique que les deux lettres C et H forment un groupe de lettres inséparables et introduisent une syllabe.

- Mais ceci pourrait être indiqué d'une manière sans doute moins évidente (deux règles), mais nécessitant au plus deux phonèmes par règle :

$$\begin{array}{l} + C + H = \left[\begin{array}{c} - \\ \cup C \cup \end{array} \right] \\ C + H + = \left[\begin{array}{c} H \cup \end{array} \right] \end{array}$$

2 Passage d'une règle formalisée à une ligne de Igram

Tout d'abord, pourquoi effectuer cette conversion caractères-entiers ? Tout simplement parce qu'il semble plus facile de manipuler des entiers que des caractères. De plus, un entier dont la valeur varie de $-(2^{15} - 1)$ à $+(2^{15} - 1)$ occupe toujours la même place mémoire : 2 octets, tandis que, pour avoir autant de possibilité différentes avec des caractères, dont chacun occupe un octet, il faudrait en regrouper ; donc, la place mémoire occupée par un groupe serait dépendante du nombre de caractères contenus dans celui-ci, ce qui rend difficile l'écriture d'un programme utilisant une telle grammaire.

Règles de conversion :

Cette conversion, et en même temps le remplissage du tableau Igram, se font au cours de l'analyse syntaxique d'une règle (cf. § 1.3 Ch. I). Celle-ci permet de distinguer les éléments suivants dans le texte d'une règle (cf. § 1-2, ch. II).

- Lettres alphabétiques
- Symboles diacritiques
- " classes "
- Symboles particuliers
- Phonèmes

Ces éléments appartiennent à différents tableaux de données :
logical * 1 alpha (56) Liste (5,9), Phon (2,47)

Voici le contenu de ces trois tableaux de caractères:

a) ALPHA

Indice	Caractère
1	[
2	E
3	i
4	Y
5	A
6	ø
7	U
8	S
9	P
10	T
11	C
12	F
13	Q
14	K
15	R
16	L
17	M
18	N
19	B
20	D

Indice	Caractère
21	G
22	H
23	J
24	V
25	Z
26	X
27	W
28	#
29	1
30	2
31	3
32	4
33	5
34	6
35	7
36	8
37	9
38	0
39	.
40	/

Indice	Caractère
41	:
42	;
43	!
44	?
45	-
46	(
47)
48	+
49	=
50	[
51]
52	⌋
53	"
54	/
55	&
56	*

Dans ce tableau, les éléments ne sont pas tous placés de façon aléatoire, mais de façon à faciliter les tests nécessaires aux programmes. De même, c'est pour cette raison qu'un même élément y figure deux fois (indice 1 et 52) :

I, indice du tableau ALPHA	caractères correspondants
$2 \leq I \leq 7$	voyelles ("VOYEL")
$8 \leq I \leq 27$	consonnes ("CONSO")
$2 \leq I \leq 27$	lettre alphabétique ("A...Z")
$2 \leq I \leq 4$	voyelles E-I-Y ("E-I-Y")
$5 \leq I \leq 7$	voyelles A-O-U ("A-O-U")
$15 \leq I \leq 16$	"liquides" ("LIKID")
$17 \leq I \leq 18$	"nasales" ("NASAL")
$8 \leq I \leq 14$ ou $17 \leq I \leq 27$	consonnes non liquides ("CNONL")
$8 \leq I \leq 16$ ou $19 \leq I \leq 27$	consonnes non nasales ("CNONN")
$1 \leq I \leq 27$	caractères pouvant figurer en 1ère position dans CT
$2 \leq I \leq 33$ et $I \neq 28$	caractères pouvant figurer dans CT, en d'autres positions

(suite du tableau)

I (indice du tableau alpha)	caractères correspondants
$1 \leq I \leq 33$	caractères pouvant figurer dans les contextes des règles.
$29 \leq I \leq 38$	caractères numériques
$39 \leq I \leq 56$	éléments séparateurs de mots dans un texte à transcrire

b) CLASSES

	1	2	3	4	5	Signification
1	C	O	N	S	O	consonnes
2	V	O	Y	E	L	voyelles
3	L	I	K	I	D	consonnes "liquides"
4	C	N	O	N	N	consonnes non "nasales"
5	A	.	.	.	Z	Lettres alphabétiques
6	N	A	S	A	L	consonnes "nasales"
7	E	-	I	-	Y	voyelles E,I,Y
8	A	-	O	-	U	voyelles A,O,U
9	C	N	O	N	L	consonnes non "liquides"

Dans ce tableau, l'ordre des lignes (1 ...9) n'a aucune importance.

c) PHON

	1	2
1	ʌ	*
2	' E	U
3	A	I
4	O	E
5	I	N
6	A	U
7	u	N
8	∅	u
9	A	N
10	∅	N
11	I	E
12	C	H
13	G	N
14	@	□
15	U	□
16	I	□

	1	2
17	E	□
18	u	□
19	A	□
20	∅	□
21	Y	□
22	W	□
23	S	□
24	Z	□
25	J	□
26	F	□
27	V	□
28	N	□
29	M	□
30	L	□
31	R	□
32	P	□

	1	2
33	B	□
34	T	□
35	D	□
36	G	□
37	K	□
38	—	□
39	Q	□
40	X	□
41	H	□
42	C	□
43	1	□
44	2	□
45	3	□
46	4	□
47	5	□

L'ordre des éléments de ce tableau n'est pas tout à fait indifférent. Les éléments ont été regroupés de la manière suivante :

I (Indice du tableau PHON)	Domaine d'utilisation
$2 \leq I \leq 37$	transcription (ordinaire)
$2 \leq I \leq 38$	transcription syllabique
$16 \leq I \leq 47$	décomposition syllabique
$2 \leq I \leq 47$	transcription libre

remarque : le premier élément de ce tableau (* *) est réservé au programme, afin de ne pas entraîner de confusion entre une transcription correcte et une éventuelle erreur.

Les éléments figurant dans ces trois tableaux différents, peuvent être distingués dans une règle, occupant une ligne du tableau Igram par :

- l'emplacement dans la ligne (cf. p.83 ch IV)
- la valeur de l'entier qui les représente.

Ceci suffit pour les phonèmes, si les entiers qui les représentent sont égaux à leur indice dans le tableau PHON : en effet, entre les indices 13 et 15 d'une ligne, ne peuvent figurer que des phonèmes. Ceci est aussi suffisant pour les caractères du terme médian de la règle, représentés par leur indice dans ALPHA.

Mais ceci ne suffit plus pour les éléments appartenant aux contextes gauche et droit d'une règle s'ils sont représentés par leur indice dans les tableaux "ALPHA" ou "LISTE". En effet, comment différencier deux éléments de même indice, l'un appartenant à "ALPHA", l'autre à "LISTE" ? De plus, comment indiquer qu'un élément de la règle, appartenant aussi à l'un ou l'autre de ces deux ensembles, est précédé du symbole "/", qui en change la signification ? (cf. chapitre. II, § 1-2-2-c).

Ces problèmes ont été résolus de la manière suivante :

- un élément de "ALPHA" est représenté par son indice dans le tableau "ALPHA".

- un élément^{de} "LISTE" est représenté par le produit de son indice dans le tableau "LISTE" et du nombre 100.

- un élément quelconque, précédé du symbole "/", est représenté par le produit de l'entier lui correspondant (cf. ci-dessus), et du nombre - 1.

ex :

Emplacement dans la ligne K de Igram :I	Valeur	élément représenté
I 13	12	PHON (1,12), PHON (2,12) = $\boxed{\emptyset}$
6 I 6 + Igram (K,12)	7	Alpha (7) = U
contexte droit 6 + Igram (K,12)I 11	24	Alpha (24) = V
contexte gauche 2 I 5	- 18	consonne non égale à (alpha(18)=N)
contexte gauche 2 I 5	400	"Liste (1,4)... Liste (5,4)" = "LIQUID" = L ou R
contexte droit 6 + Igram (K,12) I 11	- 400	Elément quelconque \neq L et \neq R

Si au cours de cette opération de codage d'une règle, une erreur est détectée dans cette dernière, celle-ci est signalée et la ligne correspondante du tableau Igram est remise à zéro.

-3 - Partitionnement et classement des règles.

Cette opération est indispensable pour un bon fonctionnement de la grammaire (cf. § 2-2, chap. II).

La partition. Les règles de la grammaire, ou encore les lignes occupées du tableau Igram, sont réparties en 27 listes disjointes, dont chacune représente une des parties de la grammaire présentées précédemment (cf. § 2-2-1, ch. II)

Le classement. Les règles ayant été converties en nombres entiers, dans le tableau Igram, le classement de celles-ci (cf. 2-2-2.c, ch. II) s'en trouve facilité : en effet, L D est alors égale au nombre d'entiers non nuls d'une ligne de Igram, situés entre la 6ème et la 11ème colonne du tableau. La L G est elle, égale au nombre d'entiers non nuls d'une ligne, situés entre la 2ème et la 5ème colonne du tableau Igram. Les différentes lignes de Igram sont alors classées en 27 parties, suivant L D décroissante et, à L D égales, suivant L G décroissante. L'ordre est indiqué par la 1ère colonne de Igram : dans chaque ligne figure un premier entier indiquant l'indice de la ligne de Igram, où se trouve la règle suivante dans le classement adopté. Si cet entier est nul, cela indique qu'il n'y a pas de règle suivante.

Ex: Codage numérique des règles correspondant à la lettre L.

80	0	28	17	3	16	16	0	0	0	0	2	30	0	0
87	0	0	28	3	16	16	0	0	0	0	2	30	0	0
78	0	0	2	3	16	28	0	0	0	0	1	11	0	0
99	0	0	0	3	16	16	0	0	0	0	2	11	0	0
100	0	0	0	0	16	16	0	0	0	0	2	30	0	0
0	0	0	0	0	16	0	0	0	0	0	1	30	0	0

Règles formalisées.

REGLE D INDICE: 79
(#+M+I)+LL+=[L]
REGLE D INDICE: 80
(#+I)+LL+=[L]
REGLE D INDICE: 87
(E+I)+L+#=[IE]
REGLE D INDICE: 78
I+LL+=[IE]
REGLE D INDICE: 99
+LL+=[L]
REGLE D INDICE: 100
+L+=[L]

Pour pouvoir utiliser ce partitionnement et ce classement, il est indispensable de connaître les indices dans Igram des règles se trouvant en tête de chacune de ces 27 listes. Pour cela, on dispose du tableau ITETE , tableau pouvant contenir 27 entiers. Là aussi, un entier nul indique qu'il n'y a pas de règle correspondante, c'est-à-dire que la liste est vide. D'autre part, ce tableau ITETE est rendu nécessaire par le fait que les entiers qu'il contient sont des variables. Ceci est dû au mode de gestion du tableau Igram. En effet, il peut au maximum contenir 320 règles, ce qui dans certains cas n'est pas excessif. Il ne s'agit donc pas de gaspiller inutilement des emplacements disponibles. Ainsi, on ne fixera pas un nombre de règles maximum pour chacune des 27 parties de la grammaire, ni un emplacement réservé aux règles "têtes de liste", qui peuvent varier.

Une règle, que l'on veut insérer dans la grammaire, sera placée au premier endroit disponible de Igram, rencontré en parcourant ce tableau à partir de la ligne 1. Une ligne K est disponible si l'un des trois éléments suivants est nul :

Igram (K,6)	Dans toute règle, le premier caractère transcrit figure nécessairement dans la 6ème colonne de Igram, représenté par son indice dans alpha (≥ 1)
Igram (K,12)	Une règle quelconque transcrit au moins un caractère : Igram (K,12) = nombre de caractères transcrits ≥ 1 .
Igram (K,13)	Une règle quelconque correspond au moins à un phonème, le premier figurant dans Igram, dans la 13 ème colonne, par son indice dans Phon (≥ 1).

Donc, pour savoir quelle est la première ligne du tableau Igram disponible pour y placer une nouvelle règle, il suffit de parcourir une colonne de Igram (6,12 ou 13) et de s'arrêter à la première ligne dont l'élément correspondant est nul.

SOUS-PROGRAMME : CODCLA (codage, classement)

De plus ce décodage des règles est indispensable car les règles de transcription codées sont pratiquement incompréhensibles à l'utilisateur. On ne peut obtenir la forme codée des règles que dans des programmes : il s'agit des "renseignements" ordre 12 (§ 4, p. 44 , ch. II).

SOUS-PROGRAMME INVERS.

6. Mise en mémoire d'une grammaire de transcription.

Afin de pouvoir choisir une grammaire de transcription pour y faire certaines modifications, ou pour transcrire un texte donné, il convient de conserver en mémoire les grammaires que l'utilisateur a mises au point. Une grammaire est donc écrite dans un fichier, dont le nom est fixé par l'utilisateur, après toute mise au point de celle-ci. Le fichier contiendra en fait, en plus d'un symbole indiquant le type d'utilisation (cf. p. 79) les deux tableaux d'entiers qui la définissent :

Itete(7) et Igram (320,15)

Le choix de la grammaire de transcription se réduira alors à la lecture d'un fichier, qui conduira au remplissage des tableaux Itete et Igram et à la mise à jour d'un symbole jouant un rôle dans la présentation du texte résultat et dans la détermination des phonèmes autorisés ou interdits dans cette grammaire.

-7 Transcription d'un mot.

(se fait au moyen de l'ordre 8 des renseignements (cf. page 44 § 4)

La grammaire de transcription est formée de nombres entiers. Il est donc logique qu'un mot, que l'on veut transcrire, soit aussi transformé en nombre entiers : ceux-ci seront les indices dans

dont le plus est "anticonstitutionnellement" comptant 25 caractères. Donc on peut en déduire que 48 caractères sont assez pour contenir deux mots entiers séparés par un caractère(␣ ou #).

Algorithme de transcription:ORTPHO(IMOT,I)

Ce sous programme est à la fois utilisé pour la transcription d'un texte, et pour la vérification des règles de la grammaire : la seule différence entre ces deux cas d'utilisation concerne les liaisons.

Paramètres :

IMOT : tableau d'entiers où est rangé le mot à transcrire.

I : indice dans IMOT du premier caractère à transcrire.

DEBUT

```
tantque(pas d'erreur)et(mot non transcrit entièrement)faire
  début
  -se placer en tête de la liste des règles correspondant à
    IMOT(I)
  -si(liste inexistante)alors ERREUR
  sinon
    début
    -tantque(une règle n'est pas validée)et(il reste au moins une
      règle dans la liste)faire
      début
      -voir si la règle est validée
      -si(règle validée)alors
        début
        -appliquer la règle
        -si(liaison)et(transcription de texte)
          alléra ECRIRE
        -mise à jour de I
        fin
      sinon passer à la règle suivante
    fin
  -si(aucune règle n'est validée)et(non liaison)alors ERREUR
  fin
fin
ECRIRE : écrire résultat
FIN
```

-8 Renseignements et modifications.

Ceux-ci découlent en fait de la structure et du mode de gestion de la grammaire, ainsi que de son mode de fonctionnement. Il n'y a donc que peu d'intérêt à s'attarder sur leur réalisation informatique.

SOUS-PROGRAMME MODIF

Conclusion.

Cette première partie a été conçue de telle sorte qu'elle puisse se suffire à elle-même, et que la grammaire qui en est le résultat, puisse être facilement mise en mémoire et utilisée par la suite pour la transcription d'un texte.

De plus, si la transcription d'un texte à l'aide d'une grammaire quelconque ne satisfait pas l'utilisateur, il lui sera possible d'obtenir les détails concernant la transcription de mots du texte à l'aide du programme MODIF, qui correspond à l'ensemble de cette première partie.

CHAPITRE V

- Exploitation de la grammaire pour une transcription quelconque d'un texte.

Pour l'une ou l'autre des "transcriptions", que les programmes permettent de réaliser, l'examen du texte se déroule suivant le même procédé ; seules les différences existant entre les grammaires utilisées permettent l'obtention d'un résultat particulier. Mais, comme la présentation du texte-résultat ne dépend pas de la grammaire utilisée, celle-ci dépendra d'autres facteurs telle la nature de la transcription. Dans la suite de ce chapitre va être présentée la réalisation informatique des principaux points de la partie commune, auxquels seront ajoutés les détails supplémentaires spécifiques à l'une ou l'autre de ces éventualités.

1 - examen du texte à transcrire

1-1 - Présentation du texte.

Le texte à transcrire est introduit de façon automatique (§1-1-b, ch. III) au cours du programme permettant de le transformer. Il serait fastidieux de le mettre "manuellement" en mémoire, à l'aide du programme d'édition implanté sur le mini-ordinateur, et qui permet la création de fichiers. En effet, chaque ligne du fichier-texte ne comporte qu'un unique caractère, ceci pour faciliter le repérage du symbole "#", indiquant la fin du texte. Cette manière de faire modifie peu le temps consacré à la lecture du fichier, car celui-ci est lu par blocs, renfermant 512 caractères au total, dont le contenu est ensuite placé en mémoire, mémoire où se feront les lectures suivantes jusqu'à épuisement du "bloc", et dont l'accès demande peu de temps.

En ce qui concerne la taille mémoire occupée sur le support par le texte (ici une disquette) elle est environ trois fois supérieure à celle occupée par un texte dont une ligne comporterait à peu près 60 caractères. Or, puisque ni la vitesse d'exécution, ni l'économie de place mémoire ne sont les buts primordiaux des programmes présentés ici, on peut admettre qu'un texte à transcrire, ne comporte qu'un caractère par ligne sur le support-mémoire où il est enregistré.

Mais ceci ne gêne en aucune manière la correction de ces fichiers par le système implanté sur le mini-ordinateur utilisé (R E D I T), car dans toutes les commandes susceptibles d'être paramétrées par une suite de caractères, le "retour chariot" est considéré comme légal.

1-2- Traitement du texte.

Un texte est en fait une suite de caractères dont seuls sont reconnus ceux figurant dans le tableau ALPHA (cf. ch. IV, p. 89), formant des mots et des nombres. Plutôt que de procéder à plusieurs transformations successives de cette suite de caractères, nécessitant pour ce faire plusieurs passages en machine d'un ensemble de caractères qui serait modifié au cours de chacun d'eux, il a semblé aussi performant de n'en effectuer qu'un seul.

En effet, bien que la plupart des programmes de transcription existants se déroulent en plusieurs étapes, il paraît possible de les réduire à une seule, non pas en rendant plus ardu la tâche de l'utilisateur, mais en modifiant et en complétant des programmes auxquels il n'a pas directement accès.

Mais , quel que soit le nombre d'"étapes" dans la transcription, la donnée de base reste la même : le texte écrit. Donc les informations dont on dispose au départ, sont exactement identiques en nombre et

en nature. Seul varie leur mode de traitement. De plus, certains passages d'un texte en machine entraînent obligatoirement d'autres :

ex : si au cours d'une étape on supprime les fins de mots muettes, il est nécessaire de procéder, dans une étape précédente, à un codage des liaisons. En effet, comment transcrire correctement : "petits enfants" à partir de "peti enfan" ?

La transcription d'un texte va donc être effectuée en une seule étape, à moins que cette transcription ne soit précédée d'une décomposition syllabique (ou segmentation graphique) des mots du texte, auquel cas deux étapes sont rendues nécessaires par l'emploi d'une grammaire unique et "spécialisée" lors d'un passage du texte en machine.

punctuation

Sont considérés comme espaces ou signes de punctuation, c'est-à-dire comme symboles séparateurs de mots et non transcrits phonétiquement, les caractères d'un texte figurant aux indices I du tableau ALPHA, tel que $39 \leq I \leq 56$ (cf. § 1.2 ch. IV). Ils permettent ainsi de marquer les frontières de mots, et d'indiquer aussi s'il y a ou non possibilité de liaison entre deux mots du texte. Seuls, les caractères alpha (52) = "␣" et alpha (45) = "-", laisseraient une possibilité de liaison ; la présence d'une autre signe de punctuation (ou considéré comme tel), l'interdit.

ex:

... petit, il ôtait ...

P Ø E T I , I L A U T A I ...

... un petit îlot ...

U N P Ø E T I T I L A U

... partit-il ...

P A R T I T I L ,

Ces signes de ponctuation, peuvent apparaître tous: (hormis l'espace et le trait d'union) à leur place correcte dans le texte transcrit ou syllabé, sur simple demande de l'utilisateur : à la console est formulée avant le début d'une transcription quelconque, la question suivante : "La ponctuation doit-elle figurer dans la transcription ? (O - N)". Cette éventualité permet de se repérer plus facilement dans le texte transcrit, mais n'est pas habituellement employée (cf. S. GSELL).

Ceci est grandement facilité par la méthode de transcription du texte, qui a lieu parallèlement à la lecture de celui-ci : donc si un élément du texte égal à alpha (I), avec $I \geq 39$ et $I \neq 45$ et $I \neq 52$, est lu, il est aussitôt inscrit dans le texte transcrit si bien sûr l'utilisateur en a fait la demande.

Il faut cependant noter que la présence, dans un texte syllabisé, de signes de ponctuation, ne gêne en rien la transcription de celui-ci : en effet, le symbole "-" qui est un trait d'union dans un texte ordinaire à transcrire, ne peut figurer en tant que tel dans le texte transcrit, que l'on décide ou non y garder les signes de ponctuation, et quelque soit la grammaire de transcription utilisée. Donc, si celle-ci permet une décomposition syllabique des mots du texte, tous les symboles "-" qui figureront dans le texte résultat indiqueront une séparation de syllabe et non pas un trait d'union. Le symbole "1" qui se trouve placé automatiquement en tête d'un texte que l'utilisateur vient d'introduire (cf. 1-1-b, ch. III), prend ici sa signification.

En effet, lorsque l'utilisateur inscrit un texte en mémoire, il est convenu que les symboles "-" qui en font partie éventuellement sont des traits d'union : c'est ce qu'indique la valeur "1" (premier passage). Si au cours de ce premier passage, la grammaire

utilisée est une grammaire permettant la décomposition syllabique des mots du texte, le texte-résultat est recopié dans un fichier dont la première ligne comporte cette fois le caractère "2". Celui-ci indiquera, au cours du traitement de ce fichier (2ème passage en machine), que tous les symboles "-" qui y figurent sont des séparateurs de syllabes dans un même mot, et non plus des signes de ponctuation séparateurs de mots. Et donc, lors de la transcription, avec cette fois-ci une grammaire ordinaire, ces symboles seront considérés comme appartenant à un mot :

ex : " homme dangereux "

UN HO-MME DAN-GE-REUX ...

UN N Ø H Ø E D AN J Ø E R EU ...

Comme on l'a vu précédemment (§1-3, chp. III), le texte est transcrit mot à mot, mais ceux-ci ne sont pas dans tous les cas des mots isolés. Pour la transcription, c'est encore le sous programme ORTPHO (cf. § 1-7, p. ch. IV) qui va être utilisé. Si, comme on l'a vu auparavant (cf. § 1-3-1, ch. III) il est possible de considérer deux mots successifs, dont seul le premier sera transcrit, ils se présenteront en forme d'une suite d'entiers rangés dans un tableau pouvant contenir 50 entiers ; ces entiers seront les indices, dans le tableau ALPHA, des caractères composants les mots .

Après transcription de ce premier mot par le sous-programme ORTPHO, transcription pour laquelle il a été tenu compte des phénomènes d'élision et de dénasalisation, étant donné que le tableau comporte deux mots, on va procéder à la transcription d'une éventuelle liaison :

Pour ce faire, la valeur 28 correspondant à la fin du mot venant d'être transcrit est transformée en 1, qui dans le codage des règles signifie une liaison. Le sous-programme ORTPHO va de nouveau être appelé, mais il sera spécifié qu'il s'agit ici d'une liaison, par la valeur donnée à une variable entière :

Ceci va permettre au sous-programme ORTPHO, de particulariser ce cas de liaison :

- le premier indice considéré dans le tableau aura la valeur K, et non la valeur 2.
- le caractère d'indice K sera seul pris en compte dans la transcription. Dans les autres cas, la transcription se poursuit, si elle peut s'achever, jusqu'à la rencontre d'une valeur 28 dans le tableau.
- si dans la grammaire de transcription aucune règle ne convient, cela ne sera pas une erreur, on considérera simplement que la liaison ne se fait pas.

Après cette étude d'un cas de liaison, le mot transcrit auparavant va être effacé et le suivant, qui était déjà dans le tableau en seconde position, va être déplacé et figurer en première position.

Avant de procéder à la transcription de ce second mot, on va définir dans le texte à transcrire s'il est (ou non) suivi d'un signe de ponctuation interdisant toute liaison, élision et dénasalisation.

Dans ce cas, le mot déjà inscrit dans le tableau sera transcrit en mot isolé puis effacé du tableau qui sera alors totalement vide.

Sinon, le mot suivant sera aussi codé en entier et inscrit dans le tableau à la suite du mot existant.

(Le cas de nombres, rencontrés dans le texte à traiter, n'est pas l'origine de modifications dans la méthode de traitement du texte; car seule est provisoirement modifiée, la provenance des mots à transcrire. (cf. § 1 -3-4, p. 52 ch. III)

PROGRAMME TRANSC

Algorithme:

```
DEBUT
N:=0;
Mettre à 0 ICARAC;
ICARAC(I)=extrémité de mots;
I:=2;
Tantque (CARLU fin de texte)faire
  début
  tantque(N > 2)et(CARLU ponctuation)faire
    début
    tantque(CARLU fin de mot)faire
      début
      ICARAC(I):=CARLU;
      I:=I+1;
      lire CARLU;
      fin;
    N:=N+1;
    sauter les blancs;
    fin;
  si(CARLU=ponctuation)alors
    début
    traitement de la ponctuation;
    sauter les blancs;
    fin
  sinon
    début
    ORTPHO(ICARAC,
    si N > 1 alors
      début
      traitement de la ponctuation;
      I:=2;
      fin
    sinon
      début
      ORTPHO(ICARAC,liaison);
      premier mot(ICARAC):=second mot(ICARAC);
      I:=fin du premier mot;
      fin;
    N=N-1;
    fin;
  fin;
FIN.
```

1-3 Particularités relatives à la nature de la transcription

a) Décomposition syllabique :

- 1° - pas d'espace laissé entre deux "phonèmes" successifs.
- 2° - Elimination automatique du symbole "-", s'il est précédé d'un symbole identique.
- 3° - Elimination automatique du symbole "-" figurant à l'extrémité d'un mot.

Ces trois points ont pour but :

- de régulariser la présentation du résultat
- de permettre le passage en machine du texte-résultat

b) transcription syllabique

1° - mots transcrits non séparés dans le texte-résultat, sauf si, dans le texte origine, ils sont séparés par un signe de ponctuation autre que l'apostrophe ou le trait d'union.

2° - Elimination automatique du symbole "-", s'il est précédé d'un symbole identique.

c) transcription libre.

Si l'on désire que le texte résultat soit sujet à un passage en machine (question formulée à la console par le programme), les "phonèmes" résultant de la transcription ne seront pas séparés par un espace.

2. Statistiques élémentaires

Celles-ci sont effectuées d'une façon très simple : pour ce faire, on dispose d'un tableau d'entiers IREG (320) dont la dimension

est égale au nombre de règles que peut contenir une grammaire. Lors de l'application de la règle I, Ireg (I) est incrémenté de 1. La connaissance des valeurs Ireg (I) permet de donner les pourcentages d'utilisation ou le nombre d'emploi de chacune des règles de la grammaire. De plus, si plusieurs textes sont transcrits avec la même grammaire, il est possible d'obtenir un résultat global sur ces textes des statistiques élémentaires, en veillant toutefois à ce que le nombre maximal qu'elles contiennent soit inférieur à $2^{15}-1$.

CONCLUSION GENERALE

En conclusion à ce travail, et bien qu'il semble répondre aux souhaits exprimés au départ, on peut citer ce qui pourrait être amélioré ou les inconvénients qu'il comporte et dont on a conscience :

a) Insertion d'une règle dans la grammaire

Bien qu'on ait essayé de construire une grammaire non ambiguë, deux cas avaient pu être signalés à l'utilisateur :

1°) Soient les deux règles :

$$R1) + e + ("conso + \#) = [AI]$$

$$R2) + e + z // = [E_{L_1}]$$

La règle R2 est un cas particulier de R1. Mais ceci n'est pas détecté par le programme qui ne s'aperçoit que de l'identité de deux règles. Ceci serait pourtant utile à l'utilisateur, et lui éviterait de corriger le classement des règles seulement après avoir constaté une erreur.

2°) Soient les deux règles :

$$R1) A + B + = [XX]$$

$$R2) + B + C = [YY]$$

Dans quel ordre classer R1 et R2 de façon à traduire correctement la chaîne graphique ABC ? Le programme ne pourra remarquer l'ambiguïté que provoque la présence simultanée dans la grammaire de R1, et de R2. Ne pourrait-il pas être modifié de façon à signaler cette ambiguïté, de façon à éviter, comme dans le cas précédent, une éventuelle correction de la grammaire.

b) Longueur maximale parfois insuffisante des contextes gauche ou droit d'une règle :

On a pu se rendre compte, surtout en ce qui concerne les liaisons, que faute de pouvoir faire figurer un nombre suffisant de caractères ou

de classes dans les contextes des règles, il pouvait en découler des erreurs, dont il est difficile de se défaire sans aucun recours à la syntaxe. Mais peut-être serait-il possible d'y remédier, en partie du moins, en n'imposant pas les longueurs maximales autorisées pour CG et (CT U CD), mais en imposant uniquement la longueur maximale autorisée pour (CG U CT U CD), tout en indiquant la taille de deux de ces trois parties d'une règle. Car, il faut noter que l'on rencontre rarement des règles dont les contextes gauche et droit sont tous deux de longueur importante.

c) Pas de notion de "choix de caractères" dans une même règle

Hormis les quelques "CLASSES" autorisées par les programmes et qui correspondent à l'UNION de certains caractères alphabétiques, l'utilisateur n'a pas la possibilité d'indiquer une possibilité de choix dans une même règle. Il doit, pour le faire, insérer autant de règles que de cas possible dans la grammaire. Ceci, bien que ne convenant pas à une grammaire ne pouvant contenir qu'un nombre limité de règles, est dû au fait qu'aucune solution n'a été trouvée pour remédier à cet inconvénient tout en autorisant un "choix" dans les règles.

exemple : + E2 + = [AI]
 + E3 + = [AI]

Comment réunir ces deux règles en une seule, dans le tableau IGRAM ?

d) Espace mémoire occupé

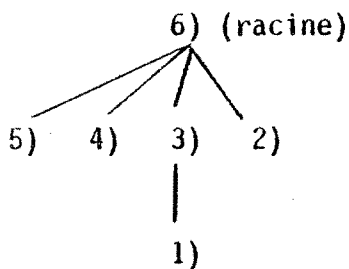
L'espace mémoire requis par ces programmes pourrait sans doute être réduit, car on n'a pas veillé à le réduire mais simplement à ne pas dépasser les limites imposées par le système. Il faut en effet remarquer que la matrice contenant une grammaire est une matrice creuse. Peut-être pourrait-on remédier à cet inconvénient ?

e) Vitesse d'exécution

Celle-ci n'a pas été préoccupante au cours de la mise au point de ce programme. Toutefois, on peut noter que les règles les plus employées, c'est-à-dire les plus générales, sont les dernières examinées et donc provoquent une perte de temps sans doute importante. Celle-ci est due au mode de classement des règles (de l'exception au cas général), et au fait

que l'on applique la première règle rencontrée qui est validée.
Ne pourrait-on pas adopter un autre mode de classement des règles, sans doute plus complexe, qui conduirait à une diminution du temps d'exécution ? Il serait possible, par exemple, au lieu de faire des listes des règles classées uniquement d'après leur longueur, de les ranger dans un "arbre", dans lequel une règle serait "fille" d'une autre règle si elle incluait celle-ci, et serait "soeur" d'une autre règle si les deux règles en question ont même mère. On pourrait alors, en commençant la recherche par la racine de l'arbre, appliquer la règle "mère" de la première "génération" de règles qui ne convient pas. Cette manière de faire réduirait le nombre de règles examinées, dans le cas des règles les plus usitées, et l'augmenterait dans des cas plus exceptionnels, sans pour autant nuire à la compréhension du fonctionnement de la grammaire :

- exemple :
- | | | | |
|----|-------|-----------|--------|
| 1) | + AIN | + "CNONN" | = [IN] |
| 2) | + AN | + "CNONN" | = [AN] |
| 3) | + AI | + | = [AI] |
| 4) | + AU | + | = [Au] |
| 5) | + A3 | + | = [Au] |
| 6) | + A | + | = [Au] |



f) Syntaxe

L'absence d'informations au sujet de la syntaxe, peut se remarquer de plusieurs manières :

- les mots homomorphes non homophones ne peuvent être distingués :

ex : fier (verbe), fier (adjectif)

portions (verbe), portions (nom)

- les liaisons ne peuvent être correctement transcrites :

ex : un avait
un_avoir

- si on poursuit ce travail en vue de la synthèse cette fois, elle est indispensable dans la prosodie.

En résumé, on pourrait peut-être dire que pour pouvoir effectuer toutes ces améliorations il aurait sûrement fallu plus de temps, un matériel plus important, etc ... Et peut-être aussi, que les circonstances extérieures (surtout le temps) ne sont pas négligeables, et que le travail effectué a essayé de s'adapter à celles-ci.

A N N E X E

CONVERSION NOMBRE - GRAPHEME

Celle-ci se fait après lecture du nombre à transcrire, à l'aide d'un sous-programme indépendant dont le rôle est d'indiquer le ou les mots qui composent ce nombre.

Voici l'algorithme de ce sous-programme : **CHIPHO (A)**, où A est un tableau de caractères contenant les N chiffres formant le nombre (N = 12).

CHIPHO (A)

- pour I:= 1 à 12 faire
 B(I):= C(I):= -1
- conversion des caractères du tableau A en entiers rangés dans le tableau B.
- pour I:= 1 jusqu'à N faire
 C(I):= N + 1 - I (rang du chiffre B(I) dans le nombre à traiter)
- transformation de C.
 Pour I:= 1 à N faire
 si (C(I)-4) div 3 ≠ 0) alors C(I):= C(I) div 3
- transformation de B.
 Pour I:= 1 à N faire
 si ((C(I) = 2) et (B(I) = 7 ou 9 ou 1)) alors
 début
 B(I):= B(I) - 1;
 B(I+1):= B(I+1) + 10;
 fin.
- remplissage du tableau d'entiers INDICE, à l'aide des tableaux B, C et de la liste des conditions présentées ci-dessous.

$C(K) = I$	$B(K) = I$	numéro condition	condition	indice dans chiffre
1	0	1	unique élément de A	31
1	0	2	condition n°1 non vérifiée	rien
1	$0 < I$	3	rien	I
2	0	4	rien	rien
2	$0 < I \leq 6$	5	élément suivant dans A égal à 1 ou 11	$I + 18$ 30
2	$0 < I \leq 6$	6	condition n°5 non vérifiée	$I + 18$
2	8	7	l'élément suivant est le dernier élément de A et est égal à zéro	32
2	8	8	condition n° 7 non vérifiée	25
0	0	9	rien	rien
0	1	10	rien	26
0	$1 < I$	11	dans A seulement par 2 éléments qui sont nuls	I 33
0	$1 < I$	12	condition n°11 non vérifiées	I 26
4	0	13	si les deux éléments éventuels le précédant dans A, sont nuls	rien
4	0	14	condition n° 13 non vérifiée	27
4	1	15	premier élément non nul de A	27
4	1	16	condition n° 15 non vérifiée	I 27
4	$1 < I$	17	rien	I 27
7	0	13	si les deux éléments éventuels le précédant dans A, sont nuls	rien
7	0	14	condition n° 13 non vérifiée	34
7	1	13	si les deux éléments éventuels le précédant dans A sont nuls	I 28
7	$0 < I$	18	$I \neq 1$ ou condition n°13 non vérifiée	I 34
10	0	13	si les deux éléments éventuels, le précédant dans A, sont nuls	rien
10	0	14	condition n° 13 non vérifiée	35
10	1	13	les deux éléments éventuels le précédant dans A sont nuls	I 29
10	$0 < I$	19	$I \neq 1$ ou condition n° 13 non vérifiée	I 35

Transcription de nombres écrits en chiffres:

S AN , ØN Z , S AN UN , D I S Y I T ,
K A R A N T Ø E Y I T , S W A S A N T Ø E D I S Y I T ,
M I L Ø E T R W A , K A T R Ø E M I L I E Ø N ,
K A T R Ø E M I L I E Ø N S I N K S A N K A R A N T Ø E
T R W A M I L Ø E S A I T S A N S W A S A N T Ø E
D I S S A I T .

TEMPS DE TRANSCRIPTION : OMN 9" 96/100

Transcription de nombres écrits en lettres:

S AN , ØN Z , S AN UN , D I S Y I T ,
K A R A N T Ø E Y I T , S W A S A N T Ø E D I S Y I T ,
M I L Ø E T R W A , K A T R Ø E M I L I E Ø N ,
K A T R Ø E M I L I E Ø N S I N K S A N K A R A N T Ø E
T R W A M I L Ø E S A I T S A N S W A S A N T Ø E
D I S S A I T .

TEMPS DE TRANSCRIPTION : OMN 5" 66/100

A N N E X E

Mode d'emploi

Modifications et renseignements constituent les seules actions possibles sur la grammaire pour sa construction et sa mise au point.
Comment peut-on les utiliser ?

Tous ceux-ci sont regroupés dans un programme qui peut être activé au moyen d'un dialogue programme - utilisateur, se déroulant à la console. Pour chacune des questions possibles présentées dans les tableaux ci-dessus, vont être décrites les différentes phases de ce dialogue : celui-ci répond aussi à l'objectif fixé, à savoir la commodité.

Pour le réaliser, il a été tenu compte de :

- l'ignorance éventuelle de l'informatique
- l'ignorance éventuelle de l'anglais.

de même que l'on a voulu éviter :

- une consultation trop fréquente des documents : constitué de questions, formulées à la console pour le programme, et de réponses que l'utilisateur devra taper sur le clavier de la console, le dialogue aura donc les caractéristiques suivantes:
 - . emploi de termes simples et courants,
 - . emploi de termes français uniquement,
 - . choix de réponses , si possible, indiqué pour chaque question,
 - . réponses courtes (souvent un seul caractère) demandées à l'utilisateur.

Ce dialogue, sous-jacent à un programme (MODIF) dont il constitue en quelque sorte un organigramme, va être décrit précisément, avec les conventions suivantes :

U - n	réponse de l'utilisateur à la n ^{ième} question posée par le programme
P - n	n ^{ième} question posée par le programme
R	résultat donné par le programme, ou action souhaitée effectuée par le programme
C	commentaire
xxxxx	contenu d'une réponse, variable

DIALOGUE

C	mise au point du programme
C	choix de la grammaire de transcription
P - 1	ancien fichier de règles (0 - N) ?
U - 1	xxxxx
C	si U - 1 est négative aller à la question P - 3
P - 2	nom du fichier :
U - 2	xxxxx

C	choix de l'ordre correspondant à la modification ou au renseignement désiré (cf. tableaux p. ,
P - 3	ordre connu (O - N) ?
U - 3	xxxxx
C	si U - 3 est affirmative aller à la question P - 4
C	visualisation des tableaux à la console afin de repérer l'ordre correspondant
P - 4	ordre choisi
U - 4	xxxxx
C	si U - 4 inférieure à 6 aller directement à la question Qu - 4
P - 5	résultats sur console ou imprimante (C - I) ?
U - 5	xxxxx
C	Q1 : suppression d'une règle
P - 6	indice de la règle
U - 6	xxxxx
C	aller à la question P - 26

C	Q2 : ajout d'une règle
P - 7	Texte de la règle
U - 7	xxxxx
C	aller à la question P - 26
C	Q3 : modification = suppression + ajout
P - 8	indice de la règle
U - 8	xxxxx
P - 9	texte de la règle
U - 9	xxxxx
C	aller à la question P - 26
C	Q4 : règle d'indice I à placer avant la règle d'indice K
P - 10	I :
U - 10	xxxxx
P - 11	K :
U - 11	xxxxx
C	aller à la question P - 26

C	Q5 : règle d'indice I à placer après la règle d'indice K
P - 12	I :
U - 12	xxxxx
P - 13	K :
U - 13	xxxxx
C	aller à la question P - 26
C	Q6 : liste classée des règles pour un caractère ou caractère par caractère (#)
P - 14	caractère désiré ou # (car. par car.)
U - 14	xxxxx
R	
C	aller à la question P - 26
C	Q7 : liste des règles comportant M caractères dans le terme médian pour un caractère donné ou caractère par caractère.
P - 15	caractère désiré ou # (car. par car.)
U - 15	xxxxx
P - 16	valeur de M (<7)
U - 16	xxxxx
R	

C	Q8 : transcription d'un mot, texte et indice des règles utilisées
P - 17	"mot" (≤ 48 caractères)
U - 17	xxxxx
R	
P - 18	on recommence (0 - N) ?
U - 18	xxxxx
C	si U - 18 est affirmative aller à la question P - 17
C	aller à la question P - 26
C	Q9 : texte d'une règle
P - 19	indice de la règle
U - 19	xxxxx
R	
P - 20	on recommence (0 - N) ?
U - 20	xxxxx
C	si U - 20 est affirmative aller à la question P - 19
C	aller à la question P - 26

C	Q10(D) : règles dont le contexte droit (gauche) comporte M termes, pour un caractère donné ou caractère par caractère
P - 21	caractère désiré ou # (car. par car.)
U - 21	xxxxxx
P - 22	nombre d'éléments dans le contexte
U - 22	xxxxxx
R	
C	aller à la question P - 26
C	Q12 : codage numérique d'une règle
P - 23	indice de la règle
U - 23	xxxxxx
R	
C	aller à la question P - 26
C	Q 13 : règles donnant même phonème
P - 25	phonème(s) dérivé(s) (< 4)
U - 25	xxxxxx
R	
C	aller à la question P - 26

C	Q 14 : règles données dans l'ordre des indices croiss.
R	
P - 26	autre modification - renseignement (O - N) ?
U - 26	xxxxx
C	si U - 26 affirmative aller à la question P - 3
P - 27	règles à ranger dans un nouveau fichier (O - N) ?
U - 27	xxxxx
P - 28	nom du fichier
U - 28	xxxxx

GLOSSAIRE

-
- DENASALISATION : suppression de la prononciation avec un son nasal; il y a dénasalisation dans bonne(B O N)par rapport au radical bon(B ON).
- DIACRITIQUE : élément rajouté à une lettre de l'alphabet pour en modifier la valeur.
ex:accents,cédille,tréma
- ELISION : chute d'un segment vocalique final devant un segment vocalique initial.
ex:l'arbre,l'image
- GEMINEE : redoublement dans l'écriture d'une lettre,n'ayant en général pas d'incidence sur la prononciation.
ex:illégal,pomme
- GRAPHEME : unité du code écrit,plus ou moins comparable à ce qu'est le phonème pour le code oral.(selon N.CATACH, le graphème est le signe substitutif du phonème;mais cette conception n'est pas partagée par tous)
- GRAPHIE : aspect visuel d'une langue telle qu'elle est représentée au moyen d'un système d'écriture.
- IDEOGRAPHIQUE : se dit d'une écriture non alphabétique,dont les éléments n'ont en général aucun rapport avec une forme phonique quelconque.
- LIAISON : prononciation,devant la voyelle initiale d'un mot, de la consonne finale du mot précédent,d'ordinaire non prononcée.
- PHONEME : plus petit segment phonique(dépourvu de sens), permettant seul ou en combinaison avec d'autres phonèmes de constituer des signifiants.
- PHONETIQUE : science qui étudie les sons d'un point de vue purement physique.(dépend de l'anatomie,de la physiologie,de l'accoustique)
- PHONOLOGIE : science qui étudie les sons d'un point de vue fonctionnel (s'oppose en cela à la phonétique)
- PROSODIE : étude de phénomènes variés, inséparables du discours tels la mélodie,l'intensité,la durée etc...
- SEMILOGIE
SEMIOTIQUE : science qui traite des systèmes et des ensembles non systématiques de signes servant à la communication.
- SYLLABATION
SYLLABISATION : découpage en syllabes d'un mot ou d'une suite de mots.

BIBLIOGRAPHIE

- AGREEN (J.), (1973), Etude sur quelques liaisons facultatives. Thèse de Doctorat, Faculté de Lettres, Uppsala.
- BENVENISTE (B.), CHARVEL (A.), (1974), L'orthographe, Maspero.
- CATACH (N.), GOLFAND (I.), DENUX (R.), (1971), Orthographe et lexicographie. Didier.
- CATACH (N.), MEISSONNIER (V.), (1979), Pour une meilleure formalisation de la conversion automatique graphème - phonème. 10e J.E.P. - G.A.L.F. - Grenoble.
- DEGRYSE (D.), GUERIN (B.), (1976), Système de transcription orthographique phonétique d'un texte français. ENSERG, Grenoble.
- DIVAY (M.), GUYOMARD (M.), (1977), Conception et réalisation d'un programme de transcription graphème - phonétique du français. Thèse de 3e Cycle, Université de Rennes.
- DIVAY (M.), GUYOMARD (M.), (1979), Le compilateur de règles de réécriture TOP et son utilisation à la transcription du français en vue de la synthèse. 10e JEP; G.A.L.F., Grenoble.
- FERVERS (H.), LE ROUX (J.), MICLET (L.), (1976), Programme de transcription orthographique phonémique en langue française, ENST, Paris.
- FOUCHE (P.), (1969), Traité de prononciation française. Klincksieck, Paris, 2e Ed.
- GAK (V.G.), (1976), L'orthographe du français, Sela, Paris.
- GALISSON (.), COSTE (.), (1971), Dictionnaire de didactique des langues. Hachette.
- GREVISSE (M.), (1964), Le bon usage. Grammaire française. Hatier, Paris - Duculot, Gembloux.
- GREVISSE (M.), (1965), Code de l'orthographe française. Baudé, Bruxelles.
- GSELL (S.), (1968), 16 Transcriptions phonétiques de textes de prose et de vers. Grenoble.
- GUYOMARD (M.), (1974), La transcription orthographique phonétique du français mise en oeuvre sur calculateurs. Rapport de stage, Lannion.
- HATON (J.P.), (1976), Bulletin du groupe Reconnaissance automatique de la parole. AFCET.
- LE CORNEC (A.), (1972), La transcription orthographique phonétique du français. Essai de formalisation. Rapport de stage, Lannion.
- LE ROUX (J.), MICLET (L.), (1979), Transcription orthographique - phonétique et synthèse en temps réel de la parole par prédiction linéaire. 10e J.E.P. - G.A.L.F., Grenoble.
- LUCCI (V.), NAZE (Y.), (1979), Enseigner ou supprimer l'orthographe ? Cedic, Paris.

- MARTINET (A.), WALTER (H.), (1973), Dictionnaire de la prononciation française dans son usage réel. France Expansion, Paris.
- MICLET (L.), ROUGEOT (B.), (1972), Transcription phonétique - orthographique de phases françaises. Note technique CEI:CSI/23.
- MOUNIN (G.), (1974), Dictionnaire de la linguistique, P.U.F., Paris.
- PRATT (B.), SYLVA (G.), (1967), Phontrs. A procedure which uses a computer for transcribing French texts into phonetic symbols. Monash University (Australie).
- PROUTS (B.), (1979), Traduction phonétique de textes écrits en français. 10e J.E.P. - G.A.L.F., Grenoble.
- SCHANE (S.), (1967), French phonology and morphology. M.I.T. Press.
- SCHANE (S.), (1967), LANGAGES : La phonologie générative. Didier-Larousse, Paris.
- TEP (G.), (1979), Système de génération des phrases phonétiques. 10e J.E.P. - G.A.L.F., Grenoble.
- THIMONNIER (R.), (1967), Le système graphique du français. Plon, Paris.
- WARNANT (L.), (1962), Dictionnaire de la prononciation française. Duculot Gembloux.