



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal & Communications



مخبر الإشارات و الإتصالات
Signal & Communications Lab.

Thèse de doctorat en Sciences

Présentée par :

El-Hadi CHERIFI

Magister en Physique Electronique

Pour l'obtention du titre de
Docteur en Électronique

Thème

*Elaboration d'un Système de Transcription Orthographique-Phonétique
en Vue de la Synthèse de la Parole à Partir du Texte en Arabe Standard*

Soutenue le 24/03/2022 devant le Jury composé de :

Mr LARBES Chérif

Mme GUERTI Mhania

Mme BENBLIDIA Nadjia

Mme HAMAMI Latifa

Mme KEDIR-TALHA Malika

Mr SAYOUD Halim

Professeur ENP-Alger

Professeur ENP-Alger

Professeur U de Blida

Professeur ENP-Alger

Professeur USTHB-Alger

Professeur USTHB-Alger

Président

Directrice de thèse

Examinatrice

Examinatrice

Examinatrice

Examineur



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal & Communications



مخبر الإشارات و الإتصالات
Signal & Communications Lab.

Thèse de doctorat en Sciences

Présentée par :

El-Hadi CHERIFI

Magister en Physique Electronique

Pour l'obtention du titre de
Docteur en Électronique

Thème

*Elaboration d'un Système de Transcription Orthographique-Phonétique
en Vue de la Synthèse de la Parole à Partir du Texte en Arabe Standard*

Soutenue le 24/03/2022 devant le Jury composé de :

Mr LARBES Chérif

Mme GUERTI Mhania

Mme BENBLIDIA Nadjia

Mme HAMAMI Latifa

Mme KEDIR-TALHA Malika

Mr SAYOUD Halim

Professeur ENP-Alger

Professeur ENP-Alger

Professeur U de Blida

Professeur ENP-Alger

Professeur USTHB-Alger

Professeur USTHB-Alger

Président

Directrice de thèse

Examinatrice

Examinatrice

Examinatrice

Examineur

ملخص: إنجاز نظام آلي لتحويل الكتابة الإملائية إلى كتابة صوتية من أجل توليف الكلام انطلاقاً من النص باللغة العربية الفصحى.

إنّ موضوع هذه الأطروحة "النسخ الإملائي-الصوتي أي تحويل الكتابة الإملائية باللغة العربية إلى كتابة صوتية". تعدّ هذه العملية مكوّنًا ضروريًا لأي نظام توليف الكلام. قمنا في بادئ الأمر بإنشاء الصياغة الصورية للإشكالات التي تميّز اللغة العربية سواء على المستوى الإملائي الخطي أو على المستوى الصوتي-الفونولوجي، ثم قمنا بعد ذلك بتطوير قواعد شاملة للنسخ الصوتي تغطّي جميع الظواهر التي تمّ تحليلها، ممّا سمح لنا بعد ذلك بتصميم نظام للتحويل الصوتي مستندا إلى تلك القواعد. قمنا في مرحلة أخيرة بإجراء اختبارات على نظامنا الذي أنجزناه، سواء على مستوى الكلمات المعزولة، أين استخدمنا في ذلك معجما صوتيا يضمّ أكثر من 35000 كلمة بتكويدها الصوتي (وفق نظام الكتابة الصوتية SAMPA)، أو على مستوى النصّ، حيث أجرينا اختبارات على نصوص مختارة بعناية تحتوي على جميع الظواهر الصوتية والفونولوجية للغة العربية مثل الغنة، وظاهرة التّمفصل الصوتي، والتّفخيم، والبلعمة وغيرها... أظهرت النتائج المتحصّل عليها أنّ المقاربة أفضت إلى دقّة أكبر من 97٪ من حيث معدّل الأخطاء في الفونيمات، ودقّة أكبر من 88٪ من حيث معدّل الأخطاء في الكلمات، ودقّة أكبر من 85٪ من حيث معدّل الأخطاء في الجمل. وكأفاق لهذه الدّراسة، سنحاول في مشروع مستقبلي، تحديد المواصفات اللازمة لدمج نظامنا للكتابة الصوتية في أنظمة توليف الكلام (TTS) الموجودة حاليا في السوق والتي تدعم اللغة العربية مثل منصّة eSpeak أو Festival أو MBROLA.

الكلمات المفتاحية: تركيب الكلام، جرافيم، فونيم، الكتابة الصوتية، الألفبائية الصوتية، الكلام المنطوق، معالجة الإشارة، اللغة العربية.

Abstract: Development of a Letter-to-sound conversion system for speech synthesis from texts in standard Arabic.

Text-To-Speech (speech synthesis or TTS) allows producing a speech signal from a given text. *The Orthographic-Phonetic Transcription (OPT)* - object of our study - is a necessary component for any TTS system. After having formulated the problems posed in the standard Arabic language (SA) both at the orthographic level and at the phonetic and phonological level, we have developed exhaustive transcription rules that cover all the analysed phenomena. Then, we proposed a rule-based OPT system that was verified using well-chosen corpus. Our system is a parameterized automaton applying a wide range of rewriting rules, which allow to associate a phoneme (or a group of phonemes) to an orthographic character (or group of characters). In this correspondence, we take into account the left context (characters or groups of characters preceding the segment to be transcribed) and the right context (characters or groups of characters following the segment to be transcribed). After having developed the formalism of OPT as well as the appropriate corresponding algorithms, we implemented these algorithms, and performing the evaluation tests carried out from a phonetic dictionary of more than 35,000 words transcribed phonetically with the SAMPA standard code. Then, the tests were operated on carefully chosen texts containing all the phonetic and phonemic phenomena of the Arabic Standard Language (AS) such as *nasalization, coarticulation, emphatization, pharyngalization...* The results show that the proposed approach gives a precision greater than 97% in terms of errors in phonemes (PER), a precision greater than 88% in terms of errors in words (WER), and a precision greater than 85% in terms of errors in sentences (SER). As a perspective for this study, we will consider in a future project, the determination of the necessary specifications for integrating of our OPT module in existing speech synthesis systems for AS, such as eSpeak, Festival or MBROLA platform.

Keywords: Speech synthesis, Grapheme, Phoneme, Transcription, Phonetic alphabet, Speech, Signal processing, Standard Arabic, Phoneme Error Rate.

Résumé : La synthèse de la parole (Text-To-Speech : TTS) permet de créer un signal de parole à partir d'un texte donné. La Transcription Orthographique-Phonétique (TOP) est un composant nécessaire de tout système TTS. Après avoir formulé les problèmes posés par la langue Arabe Standard (AS) tant au niveau orthographique qu'aux niveaux phonétique et phonologique, nous avons élaboré des règles exhaustives de transcription qui couvrent tous les phénomènes analysés. Avec de tels éléments d'analyse, nous avons proposé un système TOP à base de règles qui a été vérifié à l'aide de corpus bien choisis. Notre système est un automate paramétré appliquant une large gamme de règles de réécriture, qui permettent d'associer un phonème (ou un groupe de phonèmes) à un caractère (ou un groupe de caractères) orthographique(s), en prenant en compte le contexte gauche (caractères ou groupes de caractères précédant le segment à transcrire) et le contexte droit (caractères ou groupes de caractères suivant le segment à transcrire). Après avoir élaboré le formalisme de la TOP ainsi que les algorithmes appropriés correspondants, nous avons mis en œuvre ces algorithmes, ainsi que les tests d'évaluation qui ont été effectués à partir d'un dictionnaire phonétique de plus de 35 000 mots transcrits phonétiquement avec la norme SAMPA (Speech Assessment Methods Phonetic Alphabet). Ensuite, les tests ont été opérés sur des textes continus soigneusement choisis et renfermant tous les phénomènes phonétiques et phonémiques de l'AS telles que la *Nasalisation, coarticulation, emphatisation, pharyngalisation, etc.* Les résultats montrent que l'approche proposée donne une précision supérieure à 97% en termes d'erreurs en phonèmes (PER), une précision supérieure à 88% en termes d'erreurs en mots (WER), et une précision supérieure à 85% en termes d'erreurs en phrases (SER). Comme perspectives à cette étude, nous envisagerons dans un futur projet, la détermination des spécifications nécessaires pour l'intégration de notre module (TOP) dans les systèmes de synthèse de la parole existants pour l'AS, telles que la plateforme eSpeak, Festival ou MBROLA.

Mots clés : Synthèse de la parole, Graphème, Phonème, Transcription, Alphabet Phonétique, Parole, Traitement de Signal, Arabe Standard, Taux d'Erreur en Phonème.

Dédicaces

Je dédie cette thèse à ma chère femme Amina pour son aide et son soutien ; qu'elle trouve dans ce travail le témoignage de ma reconnaissance et de mon profond respect que je lui dois.

Mes dédicaces vont également à mes filles, Zineb El-Koubra, Hafsa Ouarda, Meriem, Imène Nesrine et Nour El-Houda. Sans oublier mes amis et collègues, à qui je ne saurai jamais être assez reconnaissant pour leurs encouragements, leur compréhension, leur tolérance et leurs prières qui ont été un énorme soutien pour ma réussite.

[oz- ɛtR ki mƏ sɔ̃ tɾɛ ʃɛR] (IPA)

Remerciements

Louange à Allah qui m'a donné force et inspiration pour réaliser cette thèse.

La paix soit sur son messager Mohammed et son honorable famille, et ses valeureux compagnons.

Tout d'abord, j'aimerais exprimer ma sincère gratitude à ma directrice de thèse, Mme GUERTI Mhania, professeur à l'ENP-Alger, pour son soutien continu dans mes études et mes recherches de Doctorat, pour sa patience et sa motivation.

Je voudrais également remercier Monsieur LARBES Chérif, professeur à l'ENP-Alger, qui a accepté de présider le jury de cette thèse.

Mes remerciements particuliers vont à Mme HAMAMI Latifa, professeur à l'ENP-Alger, à Madame KEDIR-TALHA Malika, Professeur à l'USTHB-Alger, à Monsieur SAYOUD Halim, Professeur à l'USTHB-Alger, et à Madame BENBLIDIA Nadja, Professeur à l'Université de Blida, qui, au milieu de toutes leurs préoccupations pédagogiques et académiques, ont accepté d'être membres de ce jury de thèse.

Table des matières

Liste des tableaux	
Liste des figures	
Liste des Abréviations	
Introduction générale	11
Chapitre 1 : Introduction à la production de la parole et à la synthèse de la parole <i>(Fondements, Etat de l'art et Evolution)</i>	13
1.1 Introduction	14
1.2 Processus de la synthèse de la parole	14
1.3 Evolution de la synthèse de la parole	16
1.3.1 De la synthèse mécanique à la Synthèse électrique	16
1.3.2 Développement des synthétiseurs électriques	17
1.4 Phonétique et théorie de la production de la parole	20
1.4.1 Production des sons	20
1.4.2 Modélisation de la parole	21
1.4.2.1 Représentation temps-fréquence	22
1.4.2.2 Analyse cepstrale	22
1.4.2.3 Modélisation par prédiction linéaire (LPC)	22
1.4.3 Phonétique et caractéristiques des sons de parole	23
1.4.3.1 Phonétique articuloire	23
1.4.3.2 Phonétique acoustique	24
1.4.3.3 Phonétique perceptive	24
1.4.3.4 Caractéristiques des sons	24
1.4.3.5 Phonétique et Phonologie	25
1.5 Synthèse de la parole	27
1.5.1 Définition	27
1.5.2 Principales approches de la synthèse	27
1.5.3 Traitements de la synthèse	29
1.5.3.1 Traitements linguistiques	29
1.5.3.2 Traitements numériques : Synthèse par concaténation	32
1.6 Applications de la synthèse de la parole	36
1.7 Produits TTS commercialisés sur le marché	36
1.7.1 PROJET MBROLA	37
1.7.2 GROUPE ACAPELA	37
1.7.3 ARABTALK	37
1.7.4 Sakhr TTS	37
1.7.5- Élan TTS	38
1.7.6 FESTIVAL	38
1.8 Évaluation des systèmes de synthèse de la parole	38
1.8.1 Tests subjectifs ou évaluation globale	38
1.8.2 Évaluation locale	39
1.9 Conclusion	40
Chapitre 2 : Transcription Orthographique-Phonétique en Vue de l'AS	41
2.1 Introduction	42
2.2 Rôle de la TOP dans les systèmes de Synthèse de la parole	42

2.3. Méthodes	43
2.3.1 Transcription à base de Dictionnaire	43
2.3.2 Transcription à base de règles	44
2.3.3 Transcription Data-driven	45
2.4 Problèmes de la Conversion Graphème-Phonème	47
2.4.1 Système d'écriture de l'AS	47
2.4.2 Problèmes morpho-phonémiques	51
2.4.3 Irrégularités orthographiques en AS	53
2.4.4 Noms propres et les mots étrangers	54
2.4.5 Syllabation	56
2.4.6 Normalisation du texte et pré-traitement	56
2.4.6.1 Chiffres et expressions numériques en forme indienne	59
2.4.6.2 Ligatures esthétiques	59
2.4.6.3 Ponctuation esthétique	59
2.4.6.4 Chiffres esthétiques	59
2.4.6.5 Signes de ponctuation	60
2.4.6.6 Mots-outils et harmonisations	60
2.5 Conclusion	63
Chapitre 3 : Formalisme, algorithmes, ressources numériques et Implémentation	64
3.1 Introduction	65
3.2 Formalisme du système TOP-AS	65
3.2.1 Règles phonémiques de l'AS	66
3.2.2 Règles phonétiques	72
3.2.2.1 Coarticulations anticipatoires	73
3.2.2.2 Assimilation ou Règles du « [nu:n?asa:kina] » et Nasalisation	77
3.2.3 Réalisation des règles phonologiques et phonétiques	83
3.3 Implémentation	85
3.3.1 Choix du langage de programmation	85
3.3.2 Type d'encodage/Décodage	86
3.3.3 Mise en œuvre du système TOP-AS.....	90
3.4 Ressources Numériques	92
3.5 Conclusion	102
Chapitre 4 : Test et Evaluation du système TOP-AS	103
4.1 Introduction	104
4.2 Evaluation de la TOP	104
4.2.1 Corpus de phonétisation	105
4.2.2 Evaluation de la TOP pour l'Anglais	106
4.2.3 Evaluation de la TOP pour le Français	107
4.2.4 Evaluation de la TOP pour l'Arabe Standard	108
4.3 Evaluation de notre phonétiseur TOP-AS	109
4.3.1 Corpus d'évaluation	109
4.3.2 Evaluation	122
4.3.2.1 Cas de mots isolés	122
4.3.2.2 Cas de textes	122
4.3.3 Analyse des erreurs	123
4.4 Conclusion	124
Conclusion Générale et Perspectives	125
Références Bibliographiques	128
Annexes	136

Liste des tableaux

N° Tableau	Intitulé	Page
Tableau 1.1	Exemple de symboles IPA	26
Tableau 2.1	Speech Assessment Methods Phonetic Alphabet (SAMPA) de la langue arabe (Consonnes, voyelles, diphtongues et quelques symboles).	50
Tableau 2.2	Phonèmes Consonantiques de l'AS	50
Tableau 2.3	Consonnes de l'AS et leurs correspondantes en Français (Si elles existent)	51
Tableau 2.4	Classification des consonnes tenant compte des contraintes de la transcription	52
Tableau 2.5	Supports du [hamza] au milieu du mot	53
Tableau 2.6	Supports du [hamza] à la fin du mot	53
Tableau 2.7	Liste des mots-outils et leurs harmonisations	60
Tableau 3.1	Réalisation Phonétique du [tanwi:n]	78
Tableau 3.2	Exemples de [ʔalʔidʔha:r]	78
Tableau 3.3	Exemples de [ʔalʔiqla:b]	78
Tableau 3.4	Exemples de [ʔidGa:m] avec nasalisation	79
Tableau 3.5	Exemples de [ʔidGa:m] sans nasalisation	80
Tableau 3.6	Exemples de [ʔidGa:m] total	81
Tableau 3.7	Exemples de [ʔidGa:m] partiel	81
Tableau 3.8	Les 15 « sons faibles » de l'AS	82
Tableau 3.9	Exemples de [ʔalʔiXfa:ʔ] en [nu:nʔasa:kina]	82
Tableau 3.10	Exemples de [ʔalʔiXfa:ʔ] (Assimilation 2) dans le cas du [tanwi:n]	83
Tableau 3.11 a	Standard Unicode pour les caractères de l'AS	84
Tableau 3.11 b	Codification UNICODE de l'alphabet de base de l'AS selon toutes les formes contextuelles	87
Tableau 3.12	Extrait du lexique des abréviations courantes en arabe avec leurs transcriptions correspondantes	93
Tableau 3.13	Extrait du lexique des Mots d'exception en AS avec leurs transcriptions correspondantes	96
Tableau 3.14	Extrait du lexique des Acronymes les plus répandus en langues étrangères rencontrés dans les textes arabes avec leurs sens en arabe et leurs transcriptions correspondantes.	97
Tableau 4.1	Comparaison des résultats G2P pour l'Anglais entre plusieurs sources d'étude	107
Tableau 4.2	Occurrences des règles phonologiques et phonétiques appliquées aux unités du lexique de test (Dictionnaire Phonétique MCAW-Dict).	110
Tableau 4.3	Extrait du dictionnaire phonétique de test MCAW-Dict	111
Tableau 4.4	Extrait abrégé de la liste des noms propres les plus répandus en AS	113
Tableau 4.5	Extrait du recueil de textes compilés pour les besoins d'évaluation ([su:ratulmulk] سورة الملك) avec la transcription phonétique correspondante.	114
Tableau 4.6	Extrait du recueil de textes compilés pour les besoins d'évaluation (Quelques poèmes de l'ère pré-islamique) avec la transcription phonétique correspondante.	116
Tableau 4.7	Score de la TOP sur un corpus textuel phonétisé	122

Liste des figures

N° Figure	Intitulé	Page
Fig. 1.1	Modèle de lecture spontanée	14
Fig. 1.2	Processus simple de synthèse texte -parole	15
Fig. 1.3	Modèle source-filtre de la parole.	16
Fig. 1.4	Résonateurs de Kratzenstein	16
Fig. 1.5	Reconstruction de Wheatstone de la machine parlante de Von Kempelen	17
Fig. 1.6	Synthétiseur de parole « VODER »	18
Fig. 1.7	Quelques jalons dans l’histoire de la synthèse de la parole	19
Fig. 1.8	Appareil phonatoire humain	21
Fig. 1.9	Exemple de spectrogramme (Fréquences en fonction des échantillons) d’un signal de parole échantillonné à 16 KHz	22
Fig. 1.10	Architecture d'un système de synthèse de la parole à partir du texte (synthèse par Concaténation)	30
Fig. 1.11	Représentation schématique des processus de préparation d'une Base de Données d'unités de synthèse et d'un synthétiseur de parole par concaténation.	34
Fig. 3.1	Paires des sons emphatiques et non emphatiques correspondants	74
Fig. 3.2	Modèle simplifié du conduit vocal avec déplacement de la constriction pendant la pharyngalisation d’une voyelle.	74
Fig. 3.3	Règle du [nu:nʔasa:kina] et du [tanwi:n]	77
Fig. 3.4	Architecture du système TOP-AS	91
Fig. 3.5	Séquence d’ordonnement des règles phonétiques et phonologiques du système TOP-AS	92
Fig. 4.1	Catégories d’erreurs décélées du système TOP-AS	123
Fig. 4.2	Taux d’erreurs de transcription par phonème	124

Liste des Abréviations

ANN/RNA	Artificial Neural Network/Réseau de Neurones Artificiels
API/IPA	Alphabet Phonétique International/International Phonetics Alphabet
ARMA	Auto Régressif à Moyenne Ajustée (AutoRegressive Moving Average)
AS/MSA	La langue Arabe Standard/Modern Standard Arabic
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognition/Reconnaissance Automatique de la Parole
ATTS	Arabic-TTS
BD	Base de Données
C	Consonant/Consonne
CGP	Conversion Graphème-Phonème
CRF	Conditional Random Fields / Champs Aléatoires Conditionnels
DAP	Décodage Acoustico-Phonétique
DAVO	Dynamic Analog of the VOcal tract
DCR	Degradation Category Rating
DSP	Digital Signal Processing/Traitement du signal numérique
FFT	Fast Fourier Transform
G	Gain
G2P	Grapheme-To-Phoneme
HMM	Hidden Markov Model
HNM	Harmonic plus Noise Mode
HTK	Hidden Markov Model ToolKit/boîte à outils du modèle de Markov caché
HTML	HyperText Markup Language
JMM	Joint Multi-gram Model / Modèle Multigramme Conjoint
KTH	Kungliga Tekniska Högskolan/ L'institut royal de Technologie de Suède
LAM	Laboratoire d'Acoustique
LIMSI	Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur Université Paris-Saclay
LPC	Linear Predictive Coding
LTALA	Laboratoire de Traitement Automatique de la Langue Arabe
MCAW-Dict	The Most Commonly-used Arabic Words Dictionary
MADA	Morphological Analysis and Disambiguation for Dialectal Arabic
MBROLA	Multi Band Re-synthesis by OverLap Add
MIT	Massachusetts Institute of Technology
MLP	Multi-Layer Perceptron
MLRR	Multi-Level Rewrite Rules
MPEG	Moving Picture Experts Group
NLP/TALN	Natural Language Processing/Traitement Automatique du Langage Naturel
NLTK	Natural Language ToolKit/Boîte à outils du traitement du langage naturel
OCR	Optical Character Recognition

OVE	Orator Verbis Electris
PAT	Parametric Artificial Talker
PbA	Pronunciation by Analogy
PER	Phoneme Error Rate
PSOLA	Pitch Synchronous Over Lapp and Add
RAP	Reconnaissance Automatique de la Parole
SAMPA	Speech Assessment Methods Phonetic Alphabet
SER	Sentence Error Rate
SOAP	Speech Output Assessment Package
SPC	Synthèse Par Concaténation
SPE	Sound Pattern of English
SPR	Synthèse Par Règles
TAL	Traitement Automatique des Langues
TALN	Traitement Automatique du Langage Naturel
TAP	Traitement Automatique de la Parole
TD-PSOLA	Time Domaine Pitch Synchronous Over Lapp and Add
TI	Texas Instruments
TOP	Transcription Orthographique-Phonétique
TOP-AS	Acronyme choisi pour notre système de TOP en AS
TTS	Text-To-Speech (Un Système de Synthèse à Partir du Texte)
VODER	Voice Operating DEMonstrator
WER	Word Error Rate

Introduction Générale

Les technologies de la parole se concentrent sur le développement de quatre grands types de systèmes :

- les systèmes de synthèse (Text-To-Speech) qui permettent à l'ordinateur de produire un signal vocal intelligible en imitant une voix humaine ;
- les systèmes de reconnaissance automatique de la parole (ASR) qui transforment le signal vocal en un texte lisible par ordinateur permettant ainsi d'établir une conversation bidirectionnelle ;
- des systèmes de dialogue automatique équipés d'une interface homme-machine spéciale permettant de collecter ou de fournir des informations ou d'effectuer différentes transactions homme-machine ; ce dernier système inclut les deux premiers en tant que sous-modules ;
- et les systèmes de traduction automatique de parole en parole. Ce genre systèmes sont très populaires et constituent probablement l'application la plus pertinente des technologies de la parole mentionnées ci-dessus.

Le but de la synthèse de la parole à partir du texte (ou Text-To-Speech) est de créer automatiquement un signal de parole correspondant à un énoncé écrit. Les sources du texte à prononcer peuvent être diverses : Lecture de journaux, système de réponse vocale, système d'informations, voir saisie au clavier de l'appareil. Ainsi, les deux pôles de la synthèse de la parole sont, d'un côté l'analyse et l'interprétation du texte et, de l'autre côté, la prédiction des paramètres acoustico-phonétiques du son et la synthèse finale du signal. Le premier de ces pôles entre pleinement dans le champ du Traitement Automatique des Langues (TAL) [1]. L'issue de cette étape est une séquence de phonèmes augmentés de leurs variations prosodiques. Vient ensuite l'étape de synthèse sonore proprement parler. Les sons étant définis de manière abstraite, il reste à construire l'instrument, le synthétiseur vocal qui produit la prononciation correspondante. Au début de la synthèse, jusqu'aux années 70, ce sont les questions autour des traits et des règles phonologiques qui ont constitué une partie importante de l'enjeu en synthèse [2]. Les acousticiens et les linguistes ont cherché où « se trouvait » l'invariance phonémique. Les systèmes de synthèse par règles sont apparus, avec une ambition d'universalité : systèmes multilingues, multi-locuteurs, multistyles [3]. Mais lorsqu'intervient le « locuteur » artificiel, ses caractéristiques et sa personnalité vocale si l'on peut dire, on quitte le domaine du TAL pour entrer dans celui du signal acoustique, de la synthèse sonore. C'est ainsi qu'une part importante de la recherche en synthèse de la parole n'est pas le fait de laboratoires de linguistique ni d'informatique linguistique, mais de laboratoires d'acoustique, de phonétique et de génie électrique. Enfin intervient le domaine de l'évaluation : la compréhension et la perception de la parole synthétique par les auditeurs humains.

Dans cette thèse, nous nous intéressons à la première étape des systèmes de synthèse, *et plus particulièrement à la phase de Transcription Orthographique-Phonétique (TOP)* ou encore désignée par « *La Conversion Graphème-Phonème (CGP)* ».

Un système complet de synthèse à partir du texte enchaîne d'assez nombreuses procédures de calcul [4]. Un premier ensemble de procédures, pour l'analyse et l'interprétation du texte, relève du traitement automatique de l'écrit. Tout d'abord, il s'agit de calculer la prononciation du texte. Lorsque les systèmes se sont trouvés confrontés à des sources de textes réelles et variées, il a fallu s'occuper de nombreuses « anomalies » graphémiques des textes, et de leur « normalisation » : les sigles, les abréviations, les acronymes, les diverses sortes de mots composés, les symboles, les signes et les styles de ponctuation, les divers types de nombres et de chiffres. Tout cela nécessite des prétraitements pour aboutir à une prononciation réaliste. De plus, l'apparition de textes électroniques peu structurés, voire entachés de « fautes » plus ou moins systématiques (courrier électronique) pose des problèmes nouveaux que l'on se doit de traiter automatiquement. Après cette normalisation, étape qui transforme un texte réel en suite de mots sous une forme orthographiquement correcte, il s'agit de calculer la prononciation du texte. Dès la fin des années

60, les chercheurs se sont préoccupés de cette « *orthographe inversée* » qui permet de passer des lettres aux sons, la transcription *Graphème-Phonème* » ou « *phonétisation* », et qui fait l'objet d'étude de cette thèse. Les premiers systèmes de synthèse ne comprenaient que les deux modules : le module de phonétisation (Conversion Graphème-Phonème=CGP ou TOP) et le module de synthèse acoustique de la suite de phonèmes. De tels systèmes sont d'une façon générale intelligibles, mais la parole produite n'est pas vraiment humaine, car il lui manque la prosodie : l'intonation et l'accentuation, c'est-à-dire le rythme de prononciation, les pauses dans l'énoncé, la mélodie, les variations de forces des syllabes, etc.

Nous allons nous intéresser dans cette thèse au domaine de la synthèse de la parole et plus particulièrement à l'étape *CGP*, ou encore appelée *TOP* pour la langue Arabe Standard (AS). La contribution de notre présente étude, par rapport aux travaux antérieurs sur la synthèse de la parole en langue Arabe, est l'exposition détaillée et l'analyse de tous les aspects théoriques et pratiques du problème de la conversion lettre-son en langue arabe en particulier et les autres langues en général, ce qui nous a permis la confection d'une approche à base de règles riche en outils techniques et en ressources linguistiques. Nous aborderons dans cette thèse les questions suivantes :

- les problèmes du système orthographique de la langue Arabe Standard (AS) ;
- les problèmes liés à la segmentation et au pré-traitement d'un texte en AS ;
- les différentes approches de transcription en général et pour l'AS en particulier ;
- la déduction de règles exhaustives de transcription phonétique et phonémique que nous avons adoptée pour l'AS ;
- la formulation de ces règles en algorithmes appropriés pour le traitement informatique et la mise en œuvre de ces algorithmes ;
- la Présentation des ressources utilisées (Bases de Données, corpus, lexiques, ...) et l'évaluation des résultats.

Après avoir exposé, en *chapitre 1*, l'état de l'art, l'historique, les fondements et les perspectives de la synthèse de la parole et des systèmes TTS en général, nous aborderons dans le *chapitre 2*, avec certains détails, les questions liées à la conversion du texte arabe en entités phonémiques et phonétiques pour être utilisées ensuite dans la synthèse de la parole ou dans d'autres applications. Les différentes approches de la TOP sont également étudiées en signalant les particularités de la langue arabe par rapport aux autres langues largement étudiées comme le Français et l'Anglais.

Le développement de notre formalisme de Transcription Orthographique-Phonétique à base de règles, ses algorithmes et son implémentation feront l'objet du *chapitre 3*. La conception et la mise en oeuvre de notre système de phonétisation, baptisé TOP-AS, se fait en exploitant de multitude ressources phonético-phonologiques. En effet, une période très importante de cette étude a été consacrée à étudier, collecter et annoter (transcrire phonétiquement) les ressources numériques linguistiques et phonétiques nécessaires. Ces corpus sont alignés, c'est à dire des données en AS d'un côté, et leur transcription de l'autre. Ces corpus sont exploités en partie dans l'implémentation et la construction du système de transcription automatique, puis en partie dans l'évaluation.

Enfin, les tests et l'évaluation de la qualité des résultats feront l'objet du *chapitre 4*. Une conclusion générale et des perspectives ainsi que des références bibliographiques termineront ce travail de recherche.

Chapitre 1 : Etat de l'art sur la production et la Synthèse de la Parole

1.1 Introduction

L'objectif de ce chapitre est de faire un survol des technologies de la synthèse de la parole existantes en étudiant leurs approches et techniques, leurs applications possibles ainsi que leurs méthodes d'évaluation. Nous commençons d'abord par un bref historique des différentes méthodes de synthèse de la parole et des synthétiseurs vocaux. La section suivante donne un exposé théorique rapide de la production de la parole, de la phonétique articulatoire, et d'autres concepts connexes. Cette présentation est indispensable pour une bonne compréhension du travail de cette thèse. Elle reste limitée à l'essentiel requis pour la suite. Le processus de la synthèse de la parole –dont la *Transcription Orthographique-Phonétique (TOP)* ou *phonétisation* constitue une composante décisive- engendre différents types de problèmes que nous les décrivons dans cette section. Les diverses méthodes et approches existantes pour la synthèse sont également exposées, et la section suivante est consacrée aux applications. Une liste de synthétiseurs disponibles et de certains projets de synthèse de la parole en cours sont présentés dans la *section 1.7* et enfin, la dernière section discute les méthodes d'évaluation de performance des systèmes TTS.

1.2 Processus de la synthèse de la parole

La synthèse de la parole, génération automatique de signaux de parole, a fait l'objet d'étude au cours de ces dernières décennies [5, 6]. Les progrès récents dans la synthèse de la parole ont produit des synthétiseurs de très haute intelligibilité, mais la qualité du signal de parole synthétisée et son naturel restent toujours un problème majeur. Cependant, la qualité des produits actuels a atteint un niveau suffisant pour certaines applications, telles que dans le domaine des multimédia et des télécommunications. Tout système TTS est une tentative de simulation de la lecture et de la production naturelle de la parole. La lecture spontanée se fait selon le modèle de Grainger et Ferrand [7] (*figure 1.1*).

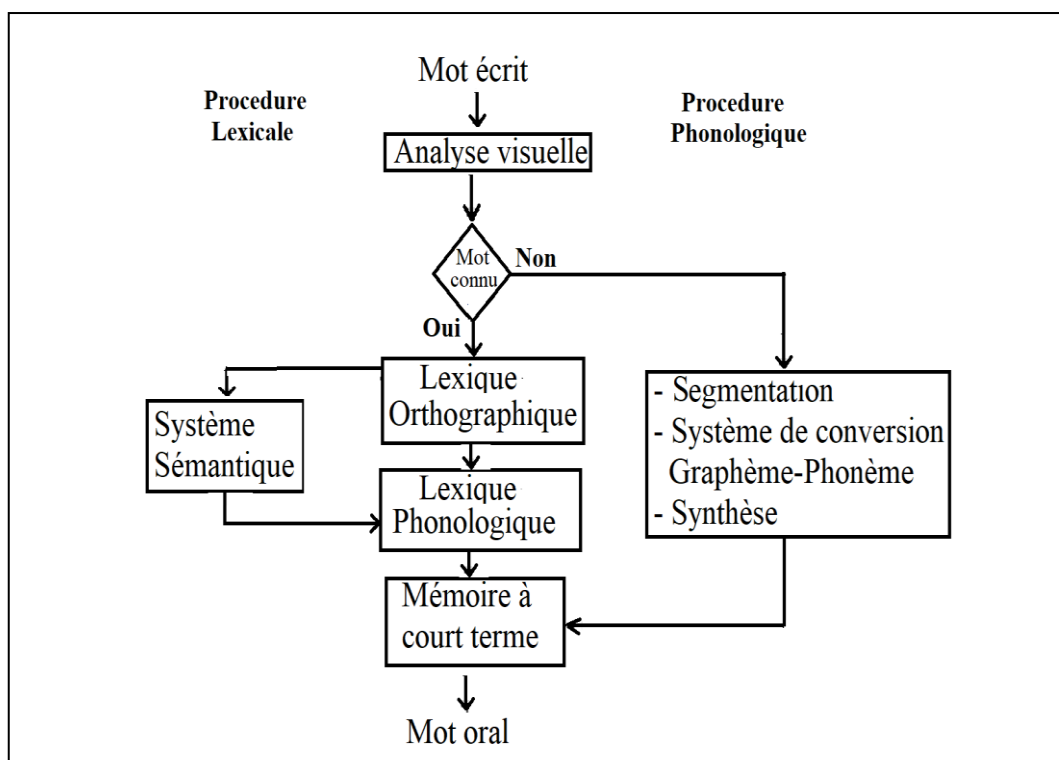


Fig. 1.1 : Modèle de lecture spontanée

Le processus de synthèse de la parole (TTS) se compose de deux phases principales. La première est l'analyse du texte, où le texte d'entrée est transcrit en une suite de phonèmes ou en une autre représentation linguistique ; et la seconde est la génération de formes d'onde de parole, où la puissance acoustique est produite à partir des informations phonétiques et prosodiques issues

de la première phase. Ces deux phases sont généralement appelées « la synthèse haut et bas niveau ». Une version simplifiée de ce processus est présentée dans la *figure 1.2*.

Le texte d'entrée peut être, par exemple des données à partir d'un traitement de texte, un ASCII standard d'un e-mail, un message textuel d'un téléphone mobile, ou un texte numérisé à partir d'un journal. La chaîne de caractères est ensuite prétraitée et analysée dans la représentation phonétique qui est habituellement une chaîne de phonèmes avec des informations supplémentaires pour une intonation, une durée et un accent correct. Les sons de la parole sont finalement générés avec le synthétiseur de bas niveau par les informations issues du synthétiseur haut niveau [8].

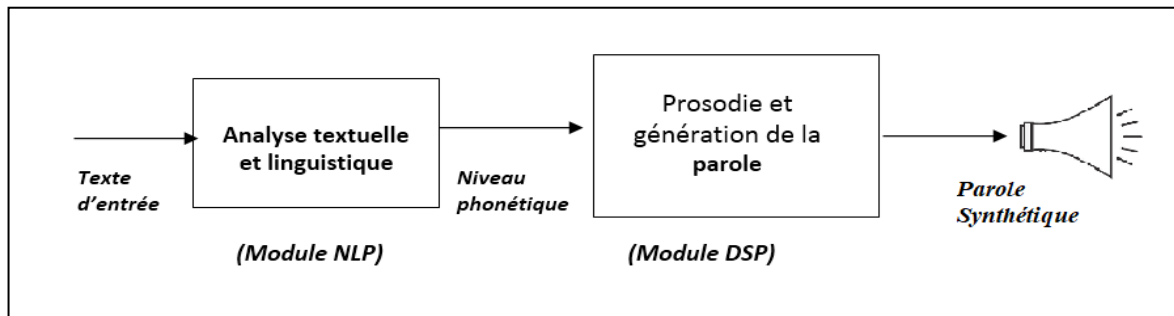


Fig. 1.2 : Processus simple de synthèse de la parole

L'approche la plus simpliste dans la génération de la parole synthétique est de reproduire de longs échantillons préenregistrés de la parole naturelle, comme des mots ou des phrases simples. Cette méthode de concaténation est de haute qualité et d'un naturel acceptable, mais elle est restreinte au cas de vocabulaire limité et généralement aux voix uniques. Cette technique est très appropriée pour certains systèmes d'annonce et d'information.

Cependant, il est tout à fait clair que nous ne pouvons pas créer une base de données acoustique de tous les mots dans une langue donnée et tous les noms communs dans le monde. Il est peut-être même inapproprié d'appeler cette technique « *synthèse de parole* », car elle est établie seulement sur les enregistrements. Ainsi, pour la synthèse de la parole libre (Text-To-Speech) nous devons utiliser des segments plus courts de signal de parole, tels que les syllabes, les phonèmes ou diphtongues. Une autre méthode largement utilisée pour produire de la parole synthétique est la synthèse à formants basée sur le modèle « *source – filtre* » de production de la parole (*figure 1.3*). La méthode est parfois appelée *analogie terminale*, car elle modélise uniquement la source du son et les fréquences des formants, et non pas toutes les caractéristiques physiques de l'appareil phonatoire humain [9]. Il est important de noter que ce modèle de production considère que la configuration du conduit vocal n'a pas d'impact sur le signal de source - d'interaction entre la source et le filtre. Ce modèle est considéré linéaire. En pratique, cette interaction existe, mais la présence d'interactions source/filtre n'empêche pas l'analyse par un modèle linéaire (LPC) s'il n'est pas trop contraint.

En théorie, la méthode la plus précise pour générer de la parole artificielle est de modéliser directement le système humain de production de la parole [10, 11, 12]. Cette méthode, appelée synthèse articulatoire, fait intervenir généralement des modèles d'articulateurs humains et des cordes vocales. Les articulateurs sont habituellement modélisés par un ensemble de fonctions d'aire de petites sections de tubes. Le modèle de la corde vocale est utilisé pour générer un signal d'excitation approprié, qui peut être par exemple un modèle à « *deux masses* » avec deux masses en déplacement vertical [13]. La Synthèse articulatoire est prometteuse pour la production d'un signal de parole de très haute qualité, mais en raison de sa complexité, ce potentiel n'est pas encore exploité.

Toutes les méthodes de synthèse précitées ont des avantages mais également des difficultés qui leur sont propres, et il semble tout à fait difficile de dire quelle méthode est la meilleure. Avec la concaténation et la synthèse à formants, des résultats très prometteurs ont été réalisés récemment, mais la synthèse articulatoire paraît aussi comme une méthode potentielle dans le futur. Les différentes méthodes de synthèse, leurs approches mathématiques et algorithmes sont discutés plus étroitement à la *section 1.5*.

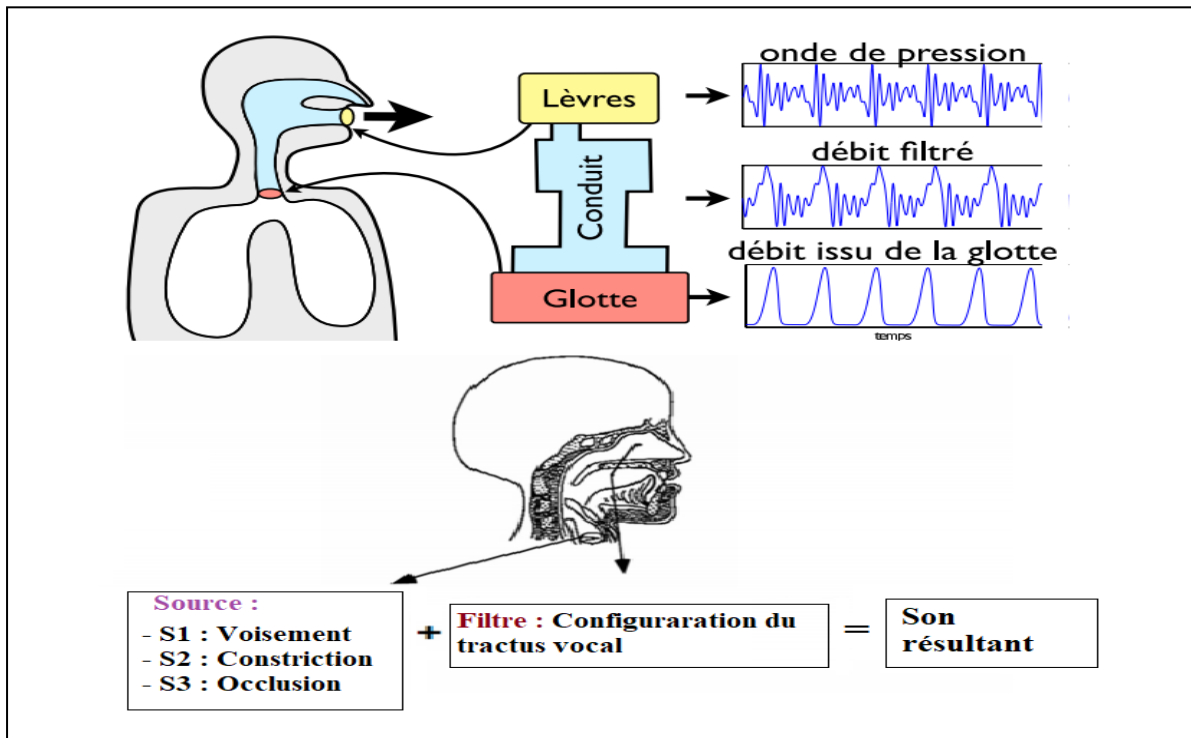


Fig. 1.3. Modèle source-filtre de la parole.

1.3 Evolution de la synthèse de la parole

Pour mieux comprendre le fonctionnement des systèmes actuels, un survol historique peut être utile. Nous discutons l'histoire de la synthèse de la parole depuis les premiers efforts mécaniques jusqu'aux synthétiseurs de haute qualité d'aujourd'hui. Pour une description plus détaillée de la parole et de l'évolution de sa synthèse, on peut se référer aux travaux exhaustifs de Klatt [14], Schroeder [15], et Flanagan [2,9].

1.3.1 De la synthèse mécanique à la Synthèse électrique

Les premiers efforts visant à produire une voix synthétique ont été faits il y a plus de deux siècles (Flanagan [2], Flanagan et al. [9], Schroeder [15]). À Saint-Petersbourg en 1779, le Professeur Russe Christian Kratzenstein étudiait les différences physiologiques entre les cinq voyelles ([a], [e], [i], [o] et [u]) et créa un appareil pour les produire artificiellement. Il a construit des résonateurs acoustiques similaires à l'appareil vocal humain et activa les résonateurs à l'aide de lames vibrantes comme avec les instruments de musique (*figure 1.4*). Le son [i] est produit en soufflant dans la conduite basse sans lame provoquant le son comme avec une flûte.

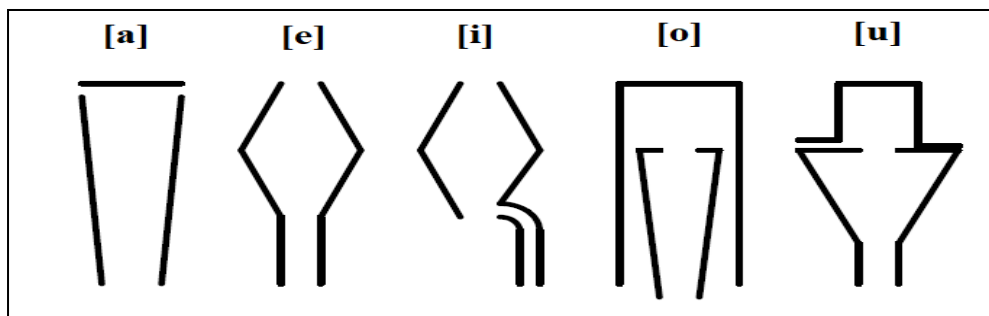


Fig. 1.4 : Résonateurs de Kratzenstein [15].

Quelques années plus tard, en 1791, Wolfgang von Kempelen a présenté à Vienne sa « *machine Acoustico-mécanique de production de parole* », capable de produire des sons simples et quelques combinaisons sonores [14, 15]. En fait, Kempelen a entamé ses recherches bien avant Kratzenstein, en 1769, et après plus de 20 années de labour, il a publié un ouvrage décrivant ses études sur la production de la parole humaine et les expériences sur sa machine parlante. Les parties essentielles de cette machine étaient une chambre de pression comme poumons, une lame vibrante comme cordes vocales, et un tube en cuir comme conduit vocal. En manipulant la forme du tube de cuir, Il pouvait produire les sons de différentes voyelles. Les consonnes ont été simulées par quatre passages séparés et contrôlés par les doigts. Pour les sons explosifs, il a également utilisé un modèle de conduit vocal qui inclut une langue articulée et des lèvres mobiles. Ses travaux ont abouti à la théorie selon laquelle le conduit vocal, c'est à dire la cavité localisée entre les cordes vocales et les lèvres, est le siège principal de l'articulation acoustique. Avant les démonstrations de Von Kempelen, le larynx a été longuement considéré comme centre de production de la parole. Au milieu des années 1800, Charles Wheatstone construit sa célèbre version de machine parlante de Von Kempelen (*figure 1.5*). Elle était un peu plus compliquée mais était capable de produire les voyelles et la plupart des consonnes. Elle pouvait produire certaines combinaisons de sons et même des mots complets.

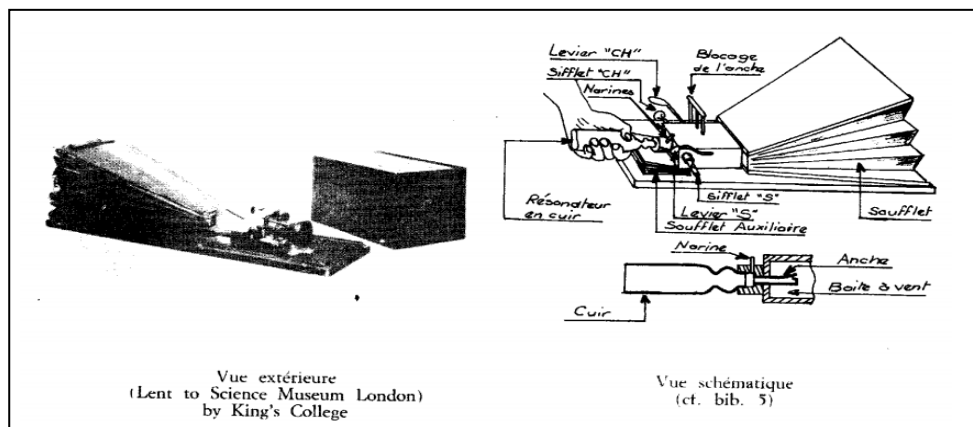


Fig. 1.5 : Reconstruction de Wheatstone de la machine parlante de Von Kempelen [16].

Le lien entre le son d'une voyelle spécifique et la géométrie du conduit vocal est découvert par Willis en 1838 [15]. Willis a également découvert que la qualité des sons des voyelles produites ne dépendait que de la longueur du tube et non de son diamètre.

À la fin du 19^{ème} siècle, Alexander Graham Bell avec son père, inspiré par la machine parlante de Wheatstone, façonna une machine parlante similaire. Les recherches et l'expérimentation sur la simulation mécanique et semi-électrique du système vocal humain ont été poursuivies jusqu'à 1960, mais sans succès remarquable. Ces travaux sont exposés en détail dans [2], et [15].

1.3.2 Développement des synthétiseurs électriques

Le premier dispositif complet de synthèse électrique a été présenté par Stewart en 1922 [14]. Le synthétiseur a un buzzer comme excitation et deux circuits résonnants pour simuler les résonances acoustiques de l'appareil vocal. La machine était capable de générer les sons simples des voyelles avec les deux formants les plus bas, mais pas de consonnes. Même genre de synthétiseur a été conçu par Wagner [9]. Le dispositif était composé de quatre résonateurs électriques connectés en parallèle, et excité par une source à buzzer. Les sorties des quatre résonateurs ont été combinées à des amplitudes ajustées pour produire des spectres de voyelles. En 1932, les chercheurs japonais Obata et Teshima ont découvert le troisième formant en voyelles [15]. Les trois premiers formants sont généralement considérés comme suffisant pour une synthèse de la parole intelligible.

Le Premier appareil considéré comme étant un synthétiseur vocale était *VODER* (Voice Operating Demonstrator) introduit par Homer Dudley à la Foire universelle de New York en 1939 [2, 9, 14]. *VODER* a été inspiré par *VOCODER* (VOICE CODER) développé aux laboratoires Bell au milieu des années trente. Le *VODER* original était un dispositif d'analyse de la parole en variant lentement les paramètres acoustiques, qui pourrait alors conduire le synthétiseur à reconstruire une approximation du signal de parole original. Il était composé d'une barre de poignet destiné à sélectionner un voisement ou la source de bruit et une pédale à pied pour le contrôle de la fréquence fondamentale. Le signal de la source est acheminé à travers dix filtres passe-bandes dont les niveaux de sortie sont contrôlés par les doigts (*Figure 1.6*). Il a fallu beaucoup d'habileté pour produire une phrase sur l'appareil. La qualité de la parole et son intelligibilité étaient loin d'être bonne, mais le potentiel de production de la parole artificielle a été bien démontré.

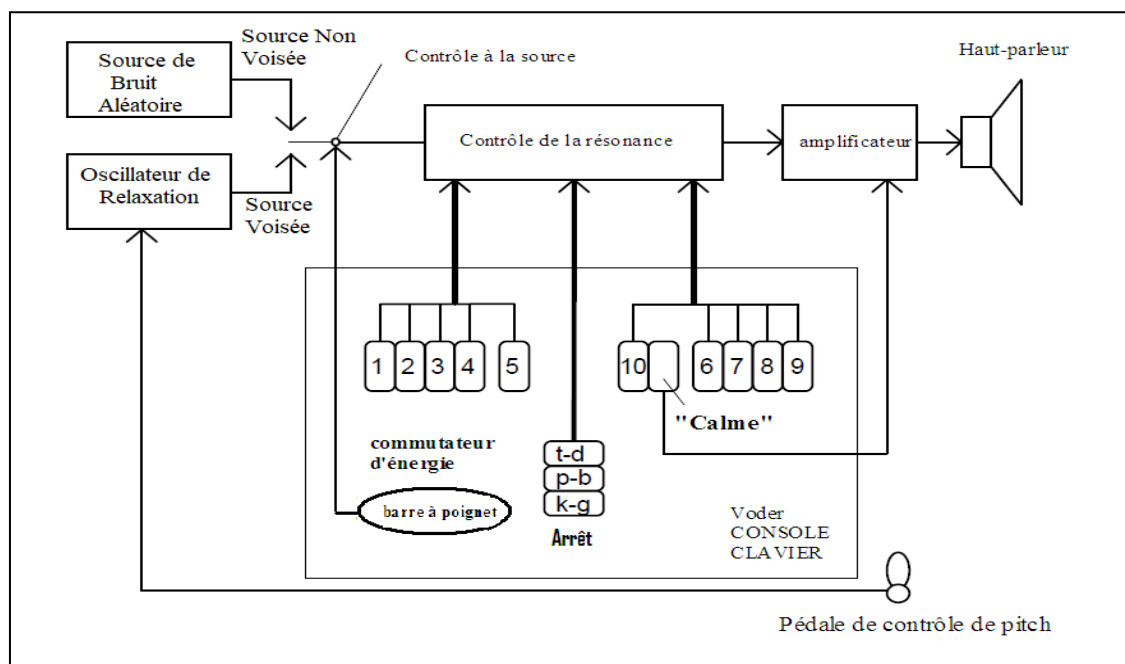


Fig. 1.6 : Synthétiseur de parole « *VODER* » Extrait de [14]

Après la démonstration de *VODER*, la communauté scientifique est devenue de plus en plus intéressée par la synthèse de la parole. Il a finalement été prouvé qu'une parole intelligible peut être produite artificiellement. En fait, la structure de base et l'idée de *VODER* a inspiré les concepteurs des systèmes actuels de production de parole basés sur le modèle source-filtre.

Près d'une décennie plus tard, en 1951, Franklin Cooper et ses collaborateurs ont développé un modèle de Synthétiseur de lecture aux Laboratoires Haskins, à l'Université de Yale en Grande Bretagne [2, 14]. Il reconvertisse les modèles de spectrogrammes enregistrés en sons, soit en leurs formes originales ou modifiées. Les modèles de spectrogrammes ont été enregistrés optiquement sur une bande transparente.

Le premier synthétiseur à formants, *PAT* (Parametric Artificial Talker), a été présenté par Walter-Laurent en 1953 [14]. *PAT* est composé de trois résonateurs électroniques à formants montés en parallèle. Au même moment, Gunnar Fant a présenté *OVE I* (Orator Verbis Electricis) le premier synthétiseur à formants en cascade, construit autour de résonateurs à formants connectés en cascade. Dix ans plus tard, en 1962, Fant et Martony introduisaient le synthétiseur *OVE II* amélioré, composé de plusieurs parties distinctes pour modéliser la fonction de transfert du conduit vocal pour les voyelles et les sons nasalisés. Ces projets *OVE* ont été suivis par *OVE III* et *GANT* au *Kungliga Tekniska Högskolan* (KTH : L'Institut royal de Technologie) en Suède, et le fameux système commercial actuel *Infovox* est leur descendant légitime [17, 18, 19].

Les synthétiseurs PAT et OVE ont été à l'origine du débat autour de la façon avec laquelle la fonction de transfert du tube acoustique doit être modélisée ; Est-ce en parallèle ou en cascade ? John Holmes a présenté en 1972 son synthétiseur parallèle à formants. Il a synthétisé la phrase [*I enjoy the simple life*] (*j'aime la vie simple*) si bien que l'auditeur moyen ne pouvait pas faire la différence entre la parole de synthèse et la phrase prononcée naturellement [14].

Le Premier synthétiseur articulatoire a été introduit en 1958 par George Rosen au Massachusetts Institute of Technology, M.I.T [14]. Le DAVO (Dynamic Analog of the VOcal tract) est contrôlé par enregistrement de signaux de commande créés à la main. Au milieu des années 1960, les premières expériences avec le Codage Prédicatif Linéaire (LPC) ont été élaborées [15]. La prédiction linéaire a été utilisée pour la première fois en 1980 dans les systèmes à faible coût, comme le TI Speak'n'Spell, mais la qualité du signal généré était assez médiocre. Cependant, avec quelques modifications apportées au modèle de base, le procédé paraît très intéressant et il est exploité dans de nombreux systèmes actuels de synthèse de la parole.

Le premier système TTS complet pour l'Anglais a été développé au laboratoire *Electrotechnical Laboratory*, au Japon en 1968 par Noriko Umeda et ses collaborateurs [14]. Il était basé sur un modèle articulatoire et incluant un module d'analyse syntaxique avec des heuristiques sophistiquées. Le signal de parole générée était tout à fait intelligible mais monotone et de qualité jugée insuffisante.

En 1979, Allen, Hunnicutt, et Klatt ont présenté le MITalk, un système TTS développé à M.I.T. Deux ans plus tard, Dennis Klatt présente son fameux système Klattalk, qui utilise une nouvelle source vocale sophistiquée décrite en détail dans [14]. La technologie utilisée dans les systèmes MITalk et Klattalk est à la base de nombreux systèmes de synthèse d'aujourd'hui, tels que DECTalk et prose-2000. La première aide à la lecture avec scanner optique a été introduite en 1976 par Kurzweil. Les machines de Kurzweil de lecture automatique pour les malvoyants étaient capables de lire assez bien un texte écrit en multi-fontes. À la fin des années 1970 et au début des années 1980, beaucoup de produits TTS commerciaux et de synthèse de la parole ont été introduits [14]. Le premier circuit intégré pour la synthèse de la parole était probablement la puce Votrax comprenant un synthétiseur à formants en cascade et de simples circuits de lissage passe-bas. En 1978, Richard Gagnon a présenté un système peu coûteux se basant sur la technologie Votrax et baptisé *Ttype-n-Talk*. Deux ans plus tard, en 1980, Texas Instruments, en exploitant la technique (LPC) introduit son synthétiseur *Speak-n-Spell* basé sur la puce de synthèse de prédiction linéaire à faible coût (TMS-5100). Il a été utilisé pour l'aide à la lecture électronique pour les enfants et reçu une attention considérable. En 1982, *Speech Plus Inc.* a introduit son système TTS *Prose 2000*. Un an plus tard, les premières versions commerciales des célèbres synthétiseurs DECTalk et Infovox SA-101 ont été introduites [14]. Quelques phases remarquables dans l'évolution de la technologie de la synthèse de la parole sont retracées dans la *figure 1.7*.

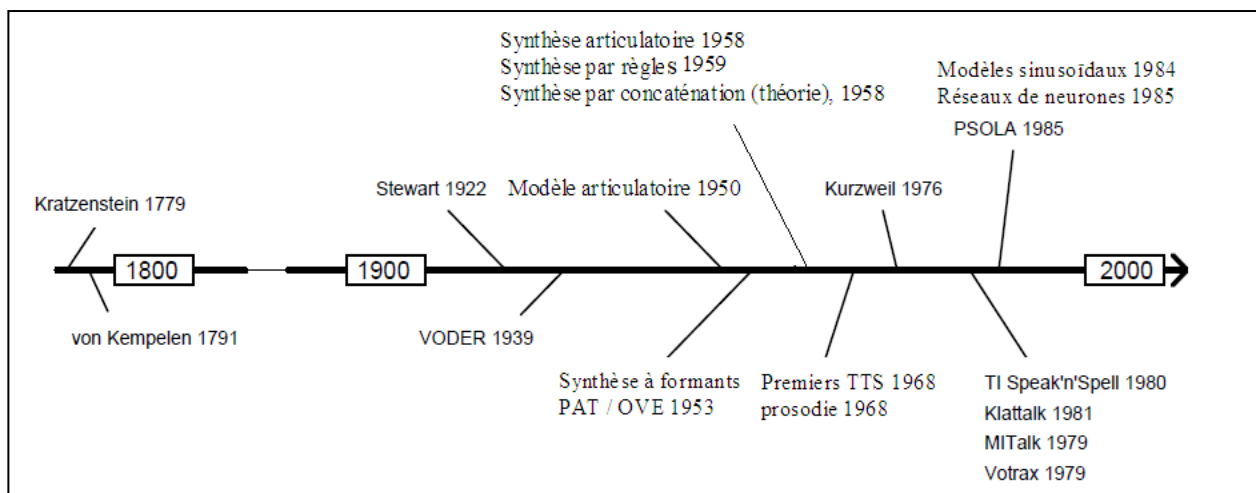


Fig. 1.7 : Quelques jalons dans l'histoire de la synthèse de la parole

Les Technologies modernes de synthèse de la parole sont fondées sur des méthodes et des algorithmes assez compliqués et sophistiqués. Une des méthodes utilisée récemment dans la synthèse de la parole est le modèle de Markov caché (HMM). Les HMM ont été appliqués à la reconnaissance de la parole à partir de la fin des années 1970. Pour les systèmes de synthèse de la parole, ils y sont exploités depuis deux décennies. Un modèle de Markov caché est un ensemble d'états reliés par des transitions avec deux ensembles de probabilités pour chacune : une probabilité de transition qui fournit la probabilité à entamer cette transition, et une fonction de densité de probabilité de sortie (notée *pdf*) qui définit la probabilité conditionnelle d'émission de chaque symbole de sortie à partir d'un alphabet fini, étant donné que la dite transition est entreprise [20]. Les Réseaux de Neurones ont été employés dans la synthèse de la parole depuis plus d'une vingtaine d'années et les derniers résultats sont très prometteurs. Cependant, le potentiel de cette approche n'est pas encore suffisamment exploré. Comme les modèles de Markov cachés, les réseaux de neurones sont également utilisés avec succès dans la reconnaissance de la parole [15].

Le premier système de synthèse automatique à partir du texte pour le Français a été conçu et réalisé au Laboratoire d'Acoustique (LAM) de Paris VI (LAM01) et au Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI-CNRS) (IM01). Ce système a été opérationnel en décembre 1968, à la suite du dépôt d'un brevet [21]. Il s'agit d'un système hybride analogique et numérique, nommé Icophone III, utilisant un ordinateur IBM 1130. La partie électronique et les concepts phonético-acoustiques ont été développés au LAM, alors que les programmes de transcription phonétique et la partie informatique ont été développés au LIMSI. La technique de synthèse est la synthèse par diphonèmes (ou diphones). Les synthétiseurs les plus célèbres (et qui intègrent la langue arabe) existant sur le marché sont exposés dans la *section 1.7*.

1.4 Phonétique et théorie de la production de la parole

Le Traitement de la parole et la technologie du langage naturel recèle beaucoup de concepts spéciaux et une terminologie diversifiée. Pour comprendre comment les différentes méthodes de synthèse vocale et d'analyse de la parole opèrent, et pouvoir également maîtriser les concepts de la phase de Transcription Orthographique-Phonétique, nous devons avoir une certaine connaissance de la production de la parole, de la phonétique articulatoire, et de la terminologie s'y rattachant. La théorie de base de ces thèmes sera discutée brièvement dans cette section.

1.4.1 Production des sons

Un son est produit lorsque l'air contenu dans les poumons est contraint à passer dans le larynx (et donc au travers des cordes vocales). En effet, l'air s'accumulant dans la cavité sub-glottique exerce une pression sur les cordes vocales et les contraint à s'ouvrir ; puis leur tension normale les aide à se refermer. L'air traverse ensuite le conduit pharyngo-buccal et nasal [22]. Des précisions sur les types de sons sont données au *paragraphe 1.4.3.4*. Sur la *figure 1.8*, les éléments les plus importants de l'appareil phonatoire qui participent à la chaîne de production d'un son sont représentés. Les vibrations quasi-périodiques des cordes vocales permettent la production de toutes les voyelles et aussi de certaines consonnes dites sonores ou voisées comme les sons [b] et [d]. La cavité buccale est un autre élément qui participe à la prononciation d'un son grâce à la mâchoire supérieure formée par les alvéoles (la partie interne de la gencive), le palais dur (au milieu) et le voile du palais. Au niveau de la partie inférieure, la mobilité de la mâchoire inférieure et la langue effectuent le déplacement. Pour la langue, on distingue la pointe le dos de la langue. Les lèvres et les dents participent aussi à la phonation.

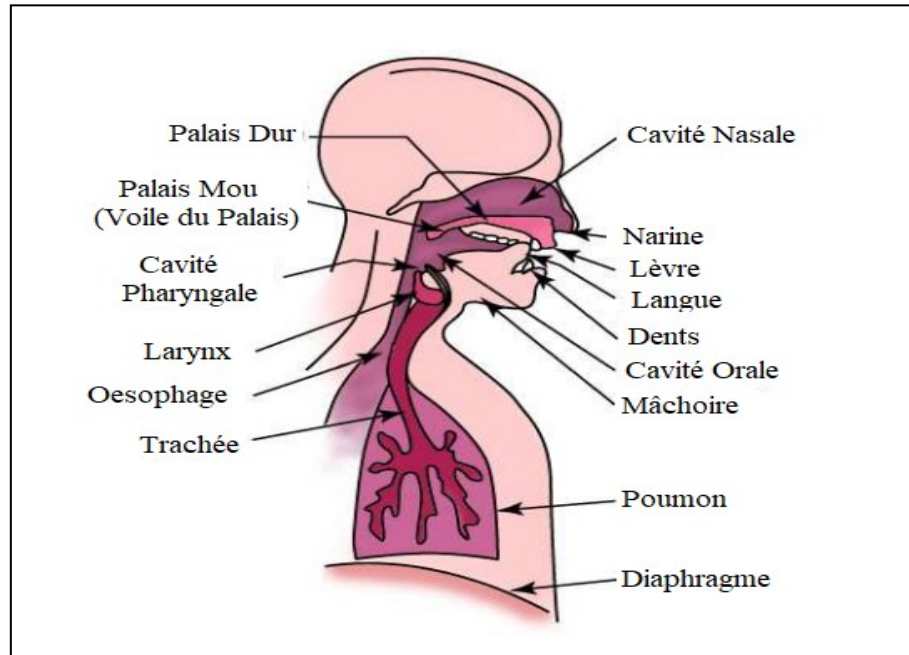


Figure 1.8 Appareil phonatoire humain [9]

1.4.2 Modélisation de la parole

D'un point de vue physique, la parole est un phénomène acoustique qui apparaît comme une variation de pression d'air causée par le système phonatoire. Les traits ou indices acoustiques d'un signal de parole sont sa fréquence fondamentale (ou pitch), son énergie, son timbre, et sa durée. Ce dernier est très lié aux traits précédents et à l'enveloppe spectrale de ce signal. Chacun de ces traits correspond à un phénomène physique ayant lieu lors de la production du son.

Le pitch représente la fréquence de vibrations des cordes vocales. Cet élément est différent pour la voix d'un homme (entre 120 Hz et 150 Hz), la voix d'une femme (aux alentours de 250 Hz) ou celle d'un enfant (entre 300 Hz et 400 Hz) [23]. L'énergie d'un son est liée à la pression de l'air en amont du larynx et caractérise son intensité. Souvent l'énergie observée dans un segment voisé est plus importante que celle observée dans un segment non-voisé. Le timbre est la caractéristique d'un son permettant de le différencier d'un autre son. Les transitoires d'attaque ou d'extinction d'un son, la présence de certains formants (enveloppe spectrale), leur renforcement ou leur atténuation sont des éléments permettant de caractériser ce timbre.

Un signal de parole est une séquence de sons correspondant à une suite d'états de l'appareil phonatoire. Le signal de parole est un processus aléatoire non stationnaire à long terme. Par contre, à court terme, il peut être considéré comme quasi-stationnaire dans le sens où son spectre de fréquence n'évolue que relativement lentement dans le temps. Cette dernière hypothèse est très utile puisqu'elle a permis de développer plusieurs techniques pour estimer ou représenter l'enveloppe spectrale. Parmi ces techniques, il y a l'analyse spectrale. Nous allons donc exposer deux représentations usuelles de l'enveloppe spectrale : la représentation temps-fréquence et le cepstre. Nous présentons également la modélisation par prédiction linéaire estimant l'enveloppe spectrale d'un signal de parole. Toutes ces méthodes utilisent l'analyse à court terme. Le traitement du signal de parole se fait donc sur des fenêtres d'analyse de l'ordre de 20 à 30 ms. Ces fenêtres se chevauchent (en général de moitié) afin d'assurer la continuité temporelle des caractéristiques de l'analyse ou du modèle et de ne pas perdre d'information comme par exemple l'explosion des plosives.

1.4.2.1 Représentation temps-fréquence

La représentation temps-fréquence est la représentation de l'évolution de l'enveloppe spectrale dans le temps. Cette représentation consiste donc à analyser en fréquence le signal de chaque fenêtre d'analyse. L'analyse fréquentielle est effectuée grâce à la Transformée de Fourier à Court Terme et nous permet de représenter le spectre de puissance d'un segment de parole. Ce spectre contient des pics appelés formants. Ces formants correspondent aux fréquences de résonance du conduit vocal [24]. Les trois premiers formants contiennent l'information la plus importante afin d'identifier les sons vocaliques.

La représentation graphique temps-fréquence d'un signal de parole est souvent faite à l'aide d'un spectrogramme. Le spectrogramme est une représentation tridimensionnelle, où le temps est représenté sur l'axe des abscisses, la fréquence sur l'axe des ordonnées et le niveau d'amplitude est symbolisé par le Niveau de Gris. Un exemple de spectrogramme d'un signal de parole échantillonné à 16 kHz est donné dans la *figure 1.9*. La représentation temps-fréquence permet de segmenter et étiqueter manuellement un signal de parole et de repérer rapidement les caractéristiques de chaque son.

1.4.2.2 Analyse cepstrale

L'analyse cepstrale permet en première approximation de séparer la contribution de la source de celle du conduit vocal, ce que ne fait pas la FFT. L'utilisation de l'analyse cepstrale est donc bien adaptée pour le calcul du pitch et des formants. Cette représentation est donc utile pour la construction de systèmes TTS à base de formants [25].

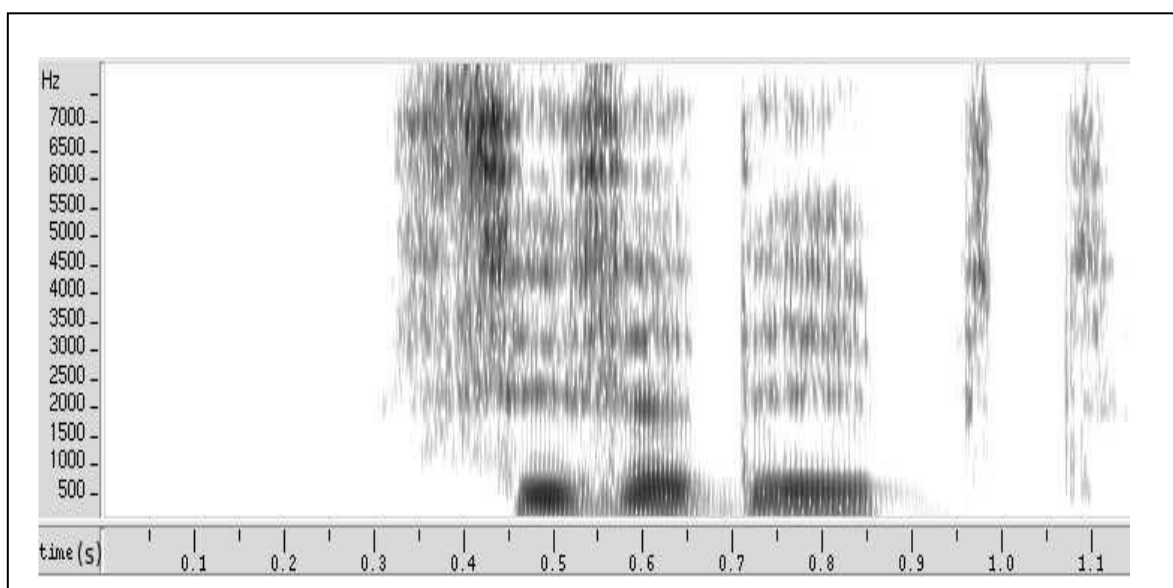


Figure 1.9 : Spectrogramme (Fréquences en fonction des échantillons) d'un signal de parole échantillonné à 16 KHz.

1.4.2.3 Modélisation par Prédiction Linéaire (LPC)

Cette technique permet de modéliser un signal de parole qui résulte de l'excitation du conduit vocal par la source glottique [26]. On peut alors modéliser les échantillons du signal de parole $s(n)$ à partir d'un signal d'excitation $e(n)$ à l'aide d'un filtre ARMA (Auto-Régressif à Moyenne Ajustée) de fonction de transfert $S(z)/E(z)$ où $S(z)$ et $E(z)$ sont les transformées en z des signaux $s(n)$ et $e(n)$. Ce modèle consiste à approcher le signal $s(n)$ par une combinaison linéaire de ses valeurs passées à laquelle on ajoute une combinaison linéaire de $e(n)$ et des échantillons précédents du signal d'excitation.

Ce modèle ARMA n'est pas simple à calculer du fait de la non linéarité des équations. On préfère recourir au modèle AR proposé par *G. Fant* en 1960 [27]. Le signal $s(n)$ est dans ce cas obtenu en ajoutant le terme d'excitation $e(n)$ (incluant l'erreur de modélisation) à une prédiction obtenue par combinaison linéaire de p échantillons précédents :

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + e(n), \quad (1.1)$$

où a_1, \dots, a_p sont les coefficients de prédiction linéaire et p est appelé l'ordre du filtre. La fonction de transfert est alors:

$$H(z) = \frac{1}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (1.2)$$

L'estimation du modèle AR revient à déterminer les coefficients a_1, \dots, a_p du filtre de fonction de transfert $H(z)$ dont on connaît le signal de sortie, mais pas complètement l'entrée. Ces coefficients sont déterminés de sorte que l'énergie de l'erreur de prédiction est minimisée dans la fenêtre d'analyse. Le modèle ainsi obtenu intègre les caractéristiques spectrales du conduit et de la source glottique. De plus, quand l'ordre de prédiction p tend vers ∞ , alors le processus $e(n)$ devient un bruit blanc. On dit dans ce cas que le filtre de fonction de transfert $H(z)$ est un filtre blanchissant. Comme en pratique l'ordre est fini, le signal résiduel $e(n)$ n'est pas blanc. Il existe de nombreuses méthodes dédiées à l'estimation des coefficients d'un modèle AR comme la résolution des équations Yule-Walker [28], l'algorithme de Levinson-Durbin, l'algorithme de Burg et la modélisation tout-pôle [29].

1.4.3 Phonétique et caractéristiques des sons de parole

La phonétique est l'étude des sons de la parole. On distingue la phonétique articulatoire qui étudie la production des différents sons, la phonétique acoustique qui étudie les propriétés physiques des sons et la phonétique perceptive qui étudie la manière dont sont perçus les sons par l'appareil auditif. Nous avons tenu à présenter cette discipline car elle est fondamentale dans la détermination des caractéristiques des sons nécessaires à la production de parole. Les aspects phonologiques fondamentaux sont également soulignés. La phonétique et la phonologie fournissent les rudiments de base pour la conception du module du processus de *Conversion Graphème-Phonème*.

1.4.3.1 Phonétique articulatoire

La phonétique articulatoire est l'étude des sons du langage humain envisagé sous l'angle de la production. Cette discipline nécessite une connaissance de la physiologie des organes de la phonation et du rôle des différents organes dans la production des sons du langage. Il est possible d'opérer une classification des sons d'une langue à partir de critères articulatoires (pour l'AS, ceci est amplement exposé au *chapitre 2*). Ces critères, au nombre de six, permettent de décrire les sons des langues et résumant, en quelque sorte, les possibilités et les limites de l'appareil phonatoire. Les critères permettant de classer les sons des langues d'une façon générale (sans tenir compte des spécificités d'une langue particulière telle que la langue arabe), sont les suivants :

- le mode articulatoire a trait à la qualité du passage de l'air dans le conduit buccal. La réalisation des voyelles implique un passage libre de l'air le long du conduit buccal. Le degré d'ouverture de la cavité buccale permet de distinguer quatre types de voyelles ; les voyelles ouvertes, les mi-ouvertes, les mi-fermées et les fermées. Pour les consonnes, deux modes articulatoires sont à distinguer. Le passage de l'air est totalement bloqué lors de la production des consonnes occlusives comme [p] [t], et [k]. Le passage est rétréci suffisamment pour permettre l'émission d'un bruit continu lors de la réalisation des consonnes fricatives comme [f] et [v] ;

- l'utilisation du conduit nasal permet de produire des voyelles ou des consonnes dites *nasales*. Lorsque le voile du palais est relevé, l'air ne passe que par la cavité buccale, donnant naissance aux sons vocaliques et consonantiques dits oraux ;
- le rôle des cordes vocales détermine le caractère sourd ou sonore des différentes articulations. Lorsque les cordes vocales vibrent, les sons seront dits *voisés* ou *sonores* par opposition aux sons *non voisés* ou *sourds* ;
- le lieu d'articulation se situe nécessairement dans la partie supérieure du conduit buccal. Les points d'articulation sont les alvéoles, le palais dur et le voile du palais. Les trois adjectifs correspondant utilisés dans les descriptions des sons sont : alvéolaire, palatal, vélaire ;
- l'articulateur est constitué par la région inférieure du conduit buccal. Il s'agit de la lèvre inférieure et des différentes parties de la langue. Les adjectifs donnés aux sons utilisant la pointe et le dos de la langue sont appelés respectivement apical et dorsal ;
- le rôle des lèvres détermine le caractère labialisé ou non labialisé d'une articulation. On distingue de la sorte les voyelles arrondies [u] des non arrondies [i] et les consonnes labialisées comme [S] des consonnes non labialisées comme [s] et [z].

1.4.3.2 Phonétique acoustique

La phonétique acoustique s'intéresse à la transmission des sons en tant que signaux acoustiques. Dans ce processus, plusieurs éléments peuvent expliquer l'origine des indices acoustiques. Nous pouvons citer entre autres, la forme du conduit vocal, l'utilisation ou non du conduit nasal, ou encore les vibrations des cordes vocales. Les paramètres acoustiques des sons les plus utilisés sont le pitch et les formants décrits précédemment. Ces paramètres acoustiques permettent de différencier des classes acoustiques de sons. Par exemple, le spectre d'un son voisé contient plus de composantes (formants) en basse fréquence qu'en haute fréquence alors que le spectre d'un son non voisé présente une accentuation vers les hautes fréquences.

1.4.3.3 Phonétique perceptive

La phonétique perceptive tente de comprendre et de décrire la perception de la parole humaine. Pour ce faire, elle a recours à la physiologie du système auditif et à la psycho-acoustique. Cette dernière est l'étude de la perception auditive en fonction du stimulus. Elle s'intéresse à la façon dont les ondes sonores sont captées par le système auditif et la manière dont elles sont interprétées par le cerveau. L'appareil auditif humain est un système très complexe mais très performant. Il se compose de trois éléments essentiels : l'oreille externe, l'oreille moyenne et l'oreille interne. Lorsqu'un son parvient au système auditif par les ondes sonores émises par la cavité buccale, il est transmis par le conduit auditif externe (oreille externe) jusqu'au tympan que les ondes font vibrer. Ensuite, l'oreille moyenne se charge de transmettre ces vibrations via les osselets jusqu'à l'oreille interne. Enfin, ces vibrations sont transformées en messages nerveux avant d'être décodés au niveau du cerveau. Une oreille normale et jeune perçoit des sons de fréquences comprises entre environ 20 Hz (sons graves) et 20000 Hz (sons aigus).

Les principales caractéristiques de la perception auditive sont l'intensité perçue des sons, la hauteur des sons, leur timbre et leur durée. L'intensité d'un son est déterminée grâce à son amplitude. Elle permet de distinguer un son fort d'un son faible et elle est mesurée en décibels. La hauteur d'un son pur ou sa tonie est déterminée par sa fréquence. Le timbre détermine la couleur d'un son. Il est différent pour chaque type de source sonore et différencie, à l'oreille, deux sons qui auraient la même fréquence fondamentale et la même intensité. Par exemple, deux sons de même intensité provenant de deux instruments différents ont des timbres différents.

1.4.3.4 Caractéristiques des sons

Cette partie décrit certaines caractéristiques acoustiques des voyelles et des consonnes. Notons que ces caractéristiques sont très utiles pour la synthèse de la parole-et plus particulièrement les deux phases clés : phonétisation et synthétisation - et la compréhension des erreurs dans de tels systèmes. Concernant la langue arabe, ce point est largement exposé dans le *chapitre 2*.

Les voyelles sont toutes des sons voisés puisqu'elles proviennent des vibrations des cordes vocales. Le timbre propre à chacune dépend principalement de la position de la langue ou de l'ouverture de la bouche. On peut ainsi distinguer des classes de voyelles selon leur lieu

d'articulation (antérieures ou postérieures), le degré d'ouverture du conduit vocal (fermées ou ouvertes), la nasalisation (nasale ou orale), la forme des lèvres (labiales ou arrondies), etc.

Du point de vue acoustique, les voyelles sont caractérisées par leurs formants F_1 , F_2 et F_3 . Plus la fréquence F_1 est grande, plus la voyelle est ouverte, tandis que plus F_2 est grande, plus la voyelle est antérieure. La connaissance de F_3 est utile pour distinguer les classes proches acoustiquement et la labialité. Les caractéristiques des consonnes sont plus complexes du fait de l'existence de différents modes d'articulation et de la présence de deux sources : une source de bruit (consonnes non voisées) et une source sonore (consonnes voisées). Avec ces critères, nous pouvons distinguer plusieurs classes de consonnes : les fricatives non voisées, les fricatives voisées, les plosives non voisées (sourdes), les plosives voisées et les sonantes.

Les plosives ou occlusives sourdes [30] se caractérisent par un silence initial provenant de la fermeture totale du conduit vocal suivi d'une explosion. Les plosives voisées sont par définition des sons voisés contenant une explosion. Les fricatives sont des bruits qui résultent d'une turbulence généralement créé par un fort resserrement du conduit vocal [27].

Les sonantes sont des sons sonores qui perdent leur voisement à côté d'une consonne sourde (non voisée). Dans ce groupe, plusieurs sous-classes peuvent être distinguées : les consonnes nasales, les liquides et les semi-voyelles. Pour différencier ces trois types de consonnes, trois indices peuvent être pris en compte : la présence et la durée du silence, la présence du bruit ainsi que la position des formants.

L'identification perceptive des sons (voyelles ou consonnes) est basée sur des indices acoustiques pertinents qui permettent de distinguer visuellement les différentes classes de sons. Pour les voyelles, ces indices correspondent à la position, la largeur de bande et l'intensité des formants. Il existe plusieurs indices discriminants pour les consonnes telles que la durée entre le relâchement et la vibration des cordes vocales. Cette durée est appelée VOT (Voice Onset Time) et elle est définie uniquement pour les plosives.

Les différentes classes de phonèmes de l'AS ainsi que leurs correspondances en Français sont exposées au *chapitre 2*.

1.4.3.5 Phonétique et phonologie

La phonétique et la phonologie sont deux branches de la linguistique qui interprètent le même matériau : la parole. La phonétique étudie les sons des langues du monde en tant que réalité physique (production, transmission et perception de ces sons), tandis que la phonologie recherche les principes qui régissent leur apparition et leur fonction de codage d'une langue particulière [31].

En fonction de la configuration géométrique du conduit vocal, différents sons sont produits (*paragraphe 1.4.3.1*). Chacun des articulateurs du conduit vocal peut en effet prendre un nombre considérable de positions. Le passage d'une position à une autre ne se fait pas de manière abrupte mais sur un continuum. Mais seulement un nombre restreint de configurations de l'appareil phonatoire sont utilisées linguistiquement pour la production des sons de parole que sont les phones. La phonologie utilise le principe de regroupement des phones en fonction de leur capacité de différenciation linguistique pour une langue donnée. Ce principe conduit à la définition de classes de phones ayant la même fonction linguistique : les phonèmes [32].

L'unité de codage linguistique de la phonologie est en effet le phonème. Selon la théorie des traits distinctifs, chaque phonème se caractérise par un ensemble de traits qui le différencie des autres phonèmes. Ces traits distinctifs reflètent des propriétés de nature acoustico-auditive ou articulaire [32, 33]. Le phonème est une unité phonologique distinctive (La substitution dans un mot d'un son phonétique par un autre caractérisé par un symbole phonémique différent change le sens de ce mot) et commutable (La substitution dans un mot d'un son phonétique par un autre caractérisé par le même symbole phonémique ne change pas le sens de ce mot). Exemples: pour le Français, la substitution de la consonne occlusive labiale sonore du mot *bain* par une consonne occlusive labiale sourde (*pain*) provoque un changement de sens. Les consonnes [b] et [p] sont donc deux phonèmes différents.

En revanche, la substitution d'une consonne [R] grasseyée (dorsale) par une consonne [r] roulée (apicale) dans un mot comme *renard* ne change pas le sens de ce mot. Les consonnes [R] et [r] sont donc deux variantes d'un unique phonème. Pour la langue arabe, l'exemple des deux mots « سيف » (épée) et « صيف » (été) révèle que « س » [s] et « ص » [s'] sont deux phonèmes différents.

La phonétique articulatoire peut donc être définie comme l'étude de l'articulation des phonèmes. Un phonème peut en effet avoir plusieurs variantes articulatoires (phonétiques), appelées les allophones. L'apparition de ces variantes est due au phénomène de coarticulation (*que nous allons détailler pour la langue arabe au chapitre 2, puis formaliser au chapitre 3*) : comme chaque articulateur du conduit vocal évolue de façon continue, les mouvements articulatoires peuvent être modifiés, suivant le contexte et sous l'effet de l'inertie mécanique de ces articulateurs, de manière à minimiser l'effort à produire pour réaliser une séquence de phonèmes donnée. Les consonnes liquides par exemple, se caractérisent par une forte sensibilité à la coarticulation d'où une grande variabilité acoustique. Les réalisations physiques d'un phonème peuvent aussi varier en fonction d'autres facteurs tels que le locuteur et le dialecte [34].

La transcription phonétique d'un énoncé de parole consiste à noter la séquence des symboles représentant la séquence des sons phonétiques constituant cet énoncé. La représentation orthographique d'un énoncé n'est pas en effet suffisamment indicatrice de sa prononciation phonétique, surtout pour certaines langues comme l'Anglais et le Français (Exemple : il a eu un peu... pour le Français, et : *Enough* qui se prononce /iynahf/ pour l'Anglais). En AS, ce phénomène de décalage entre la représentation orthographique et la prononciation est tolérable. La Transcription Phonétique doit utiliser des symboles distincts pour tous les sons qui assurent des distinctions linguistiques. Mais la finesse de la notation peut être beaucoup plus grande en recensant les allophones des différents phonèmes.

L'API propose un alphabet phonétique qui associe des symboles aux sons phonétiques de la parole (définis par une description articulatoire) de façon à permettre la notation des prononciations des mots de chaque langue (*Tableau 1.1*).

Tableau 1.1 : Exemple de symboles IPA, extrait du site : [IPA English charts \(internationalphoneticassociation.org\)](http://IPA-English-charts.internationalphoneticassociation.org) révisé en 2020)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Les symboles à droite dans une cellule sont pour les voisées, à gauche sont pour les sourdes. Les zones grisées désignent des articulations jugées impossibles.

1.5 Synthèse de la parole

La synthèse de la parole est l'une des technologies les plus importantes de ces dernières années en raison de l'expansion du champ d'applications. Plusieurs travaux sur les systèmes TTS ont été réalisés pour l'Anglais et le Français, alors que de nombreuses autres langues, dont l'AS, ont été récemment prises en considération. Le domaine de la synthèse de la Parole en AS n'a pas suffisamment progressé et il en est encore à sa première étape avec une faible qualité du signal de la parole synthétisé.

1.5.1 Définition

L'objectif de la synthèse de la parole est de produire une voix intelligible et naturelle à partir d'une représentation phonétique d'un texte donné [35]. Pour ce faire, elle doit gérer toutes les coarticulations entre les sons et générer la prosodie de chaque son. Le conduit vocal, comme tout système physique, possède une certaine inertie qui l'empêche de "passer" instantanément d'un phone à un autre. La modification de la configuration du conduit vocal se fait donc de façon progressive. Les deux phones subissent alors une distorsion appelée coarticulation. Un phone (allophone) est la plus petite unité acoustique qui différencie un discours. En fait, chaque unité abstraite est appelé phonème, et même si une langue a un ensemble fixe de phonèmes, leur prononciation dépend notamment des effets contextuels, des caractéristiques vocales individuelles du locuteur, son humeur et ses intentions. La représentation acoustique d'un phonème est appelé phone. En parole continue, la prononciation de chaque phone est influencée par l'articulation des phones voisins. Les variations de prononciation dues au contexte des phones sont représentées par les allophones. Les allophones sont donc les différentes performances pour le même phonème, et donc sont prononcées avec des configurations différentes de l'appareil vocal.

La prosodie permet d'exprimer les effets sonores de l'énoncé tels que l'intonation, l'exclamation, le rythme et l'accentuation. Les informations données par la prosodie se manifestent par des variations de fréquence, de durée et d'intensité [36].

Les systèmes de synthèse qui nous intéressent sont des systèmes TTS qui prennent en entrée un énoncé sous forme d'un texte. Ce genre de système est une machine capable de prononcer un texte à voix haute. Ce texte peut être directement généré par un clavier ou scanné, puis reconnu par un système de reconnaissance ou produit automatiquement par un système de dialogue Homme-Machine. Les systèmes de synthèse de la parole les plus connus synthétisent du texte provenant d'un clavier.

1.5.2 Principales approches de la synthèse

Au cours des deux dernières décennies, les approches de synthèse les plus intéressantes qui ont été développées sont : la synthèse par règles, la synthèse articulatoire et la synthèse par concaténation.

La synthèse par règles nécessite de connaître les mécanismes de production et de perception de la parole [37]. Le signal acoustique est d'abord analysé pour extraire une représentation simplifiée du phonème ou de ses allophones sous forme de valeurs cibles. La transition entre les valeurs cibles est ensuite modélisée à l'aide de règles contextuelles. À partir de l'ensemble des valeurs cibles et des règles de transition, un spectrogramme artificiel est généré puis on utilise des résonateurs et générateurs électriques afin de produire le signal vocal correspondant à ce spectrogramme. La technique de la synthèse par formants est de loin la plus utilisée en synthèse par règles [38]. Le but de cette technique est de générer un signal de synthèse à partir des informations sur les formants (fréquences centrales, amplitudes, largeurs de bande) et les règles d'évolution des formants entre phonèmes. Cette approche de synthèse a l'avantage de ne stocker que très peu de données mais doit recourir à un nombre très important de règles qui doivent être pertinentes afin de modéliser au mieux le signal de parole. La synthèse dite "articulatoire" se distingue de la synthèse par règles par rapport à l'élément étudié. Alors que la première tente de

générer un signal de parole en reproduisant son spectre, cette synthèse s'appuie sur une simulation de l'appareil de production, en modélisant la source d'excitation, les cordes vocales et les différents articulateurs participant à la production [39].

La synthèse par concaténation consiste à générer un signal synthétique en concaténant des unités acoustiques. Ces dernières sont obtenues par segmentation des signaux de parole. Parmi ces trois approches, celle-ci est la seule qui a permis l'émergence de systèmes de synthèse dont la qualité peut être considérée comme acceptable. Historiquement, les travaux en synthèse par concaténation ont défini un ensemble minimal d'unités permettant de vocaliser n'importe quel texte dans une langue donnée [40].

La première idée testée était d'utiliser le phonème pour plusieurs raisons : le phonème est une unité de base, bien connue des phonéticiens et quelle que soit la langue étudiée, il possède une cardinalité faible. Néanmoins, des tests ont montré la présence des discontinuités entre les phonèmes à cause du phénomène de coarticulation. Ceci a conduit à l'approche dite de synthèse par diphtonges consistant à mettre bout à bout les segments de diphtonges puis à effectuer un lissage pour diminuer les discontinuités aux points de concaténation [41]. Ces diphtonges sont préalablement enregistrés en contexte neutre, c'est-à-dire en utilisant des logatomes. Un logatome est un son vocal dépourvu de sens et composé d'une consonne (ou d'un groupe de consonnes), suivi d'une voyelle, et qui se termine par une consonne (ou un groupe de consonnes) (CVC).

Les diphtonges sont les unités acoustiques qui s'étendent du milieu de la partie stable du phonème au milieu de la partie stable du phonème suivant en tenant compte de la transition. Ensuite, des modifications prosodiques sont effectuées afin de s'approcher de la bonne prosodie, de diminuer les discontinuités et d'améliorer la qualité de la parole en sortie du système. Ces modifications peuvent être des ajustements de la longueur des diphtonges et de leur intonation par exemple grâce à la méthode TD-PSOLA qui a été mise au point par France Télécom [42]. Cette méthode permet de changer le pitch et la durée des segments à concaténer en décomposant temporellement un signal à l'aide de fenêtres recouvrantes et synchrones du pitch, puis en les reconstituant avec de nouvelles valeurs de périodes.

Une technique très récente de synthèse par concaténation est la Synthèse de ma Parole par Corpus (SPC), appelée aussi synthèse par sélection d'unités. Le principe de cette technique est de découper le corpus en unités, de sélectionner des unités de parole de taille variable dans un grand corpus et de les concaténer pour générer le signal synthétique. Le corpus utilisé, préalablement étiqueté et segmenté, offre pour chacune des unités acoustiques plusieurs contextes phonétiques et prosodiques. La phase de sélection consiste donc à choisir parmi ces contextes le candidat dont le contexte phonétique et la prosodie sont les plus proches de l'unité de la séquence phonétique et de la prosodie à synthétiser [43, 41]. Des modifications prosodiques, avec TD-PSOLA par exemple, peuvent ensuite être effectuées sur ces unités sélectionnées avant de les concaténer. Les unités acoustiques généralement utilisées sont les diphtonges mais d'autres unités acoustiques peuvent être considérées. En effet, l'utilisation d'unités plus longues diminue la densité des points de concaténation et peut produire ainsi une meilleure qualité de la parole. De nouvelles unités acoustiques qui regroupent le phone courant avec le phone qui le précède et le phone qui le suit, appelées triphones, ont été récemment utilisés dans ce type de synthèse de la parole, nécessitant de plus grandes bases de données et des méthodes de sélection plus efficaces. La Synthèse par Corpus est la technique qui produit la parole la plus naturelle mais requiert de très grandes bases de données dans la plupart des systèmes commerciaux actuels. Il arrive qu'une séquence d'unités acoustiques correspondant à une partie du texte à synthétiser soit présente dans la base de données. Sélectionner cette séquence permet d'éviter les modifications des unités de synthèse et des concaténations inutiles, ce qui préserve le naturel de la parole synthétisée ainsi produite.

Dans la *figure 1.10*, nous présentons l'architecture d'un système de synthèse. C'est une architecture modulaire qui se décompose classiquement en deux blocs de traitements séquentiels. Le texte fourni en entrée du système est analysé par un premier bloc de traitements linguistiques. Les informations issues des traitements linguistiques (séquence phonétique, les paramètres prosodiques) sont ensuite transmises au bloc de synthèse du signal de parole à proprement parler.

1.5.3 Traitements en vue de la synthèse de la parole

La procédure de synthèse enchaîne plusieurs opérations de pré-traitement avant la phase de production du signal vocal.

1.5.3.1 Traitements linguistiques

Le bloc des traitements linguistiques regroupe les fonctions suivantes :

- *Pré-traitement du texte* : l'objectif de ce traitement est la désambiguïsation du texte en entrée du système de synthèse de parole, de façon à ce que le texte pré-traité ne comporte aucune ambiguïté pour les traitements linguistiques ultérieurs. Cette étape consiste à traiter les abréviations, les acronymes, les chiffres (les numéros de téléphone et les dates par exemple), les monnaies et tous les autres symboles utilisés dans les textes [44], pour permettre la segmentation du texte en phrases et en mots ;
- *Analyse morphosyntaxique* : cette analyse comporte plusieurs étapes de traitements linguistiques successifs. Le premier traitement a pour but de séparer dans le texte pré-traité les mots grammaticaux (déterminants, pronoms, prépositions, conjonctions) des mots lexicaux. Il est indispensable en effet d'identifier les mots grammaticaux d'une phrase, car ces mots forment le squelette syntaxique de cette dernière. Le lexique des mots grammaticaux comporte généralement, outre ces mots, la prononciation phonétique de ces derniers. Les mots lexicaux sont décomposés morphologiquement en préfixes, racines et suffixes. L'analyse morphologique est d'un grand intérêt parce qu'elle permet de réduire la taille des lexiques de prononciation et elle apporte des informations sur les mots constituant une phrase : catégorie syntaxique, genre, nombre, temps, personne, etc. À l'issue de l'analyse morphologique, plusieurs mots peuvent s'avérer ambigus. Certains peuvent avoir plusieurs catégories grammaticales possibles (nom/verbe par exemple) et d'autres plusieurs prononciations phonétiques possibles (homographes hétérophones). Ces ambiguïtés peuvent impliquer une indécision sur la prononciation phonétique, et impliquent sûrement une indécision sur l'organisation de la phrase en syntagmes et en groupes de mots, et par conséquent sur la prédiction de la prosodie. Afin d'accélérer la résolution de ces ambiguïtés, on met en œuvre un analyseur contextuel dont le but est d'éliminer certaines catégories grammaticales possibles pour chaque mot, en fonction des catégories grammaticales possibles des mots voisins. L'analyse syntaxique (structurelle) intervient ensuite pour établir la catégorie grammaticale définitive de chaque mot de la phrase, ainsi que l'organisation hiérarchique de la phrase en propositions et groupes de mots et les relations de dépendance fonctionnelle entre ces groupes [45]. Cette analyse est essentielle pour les systèmes de synthèse de parole à partir du texte, notamment pour la prédiction de la prosodie qui peut utiliser cette analyse pour générer une structure prosodique de la phrase.

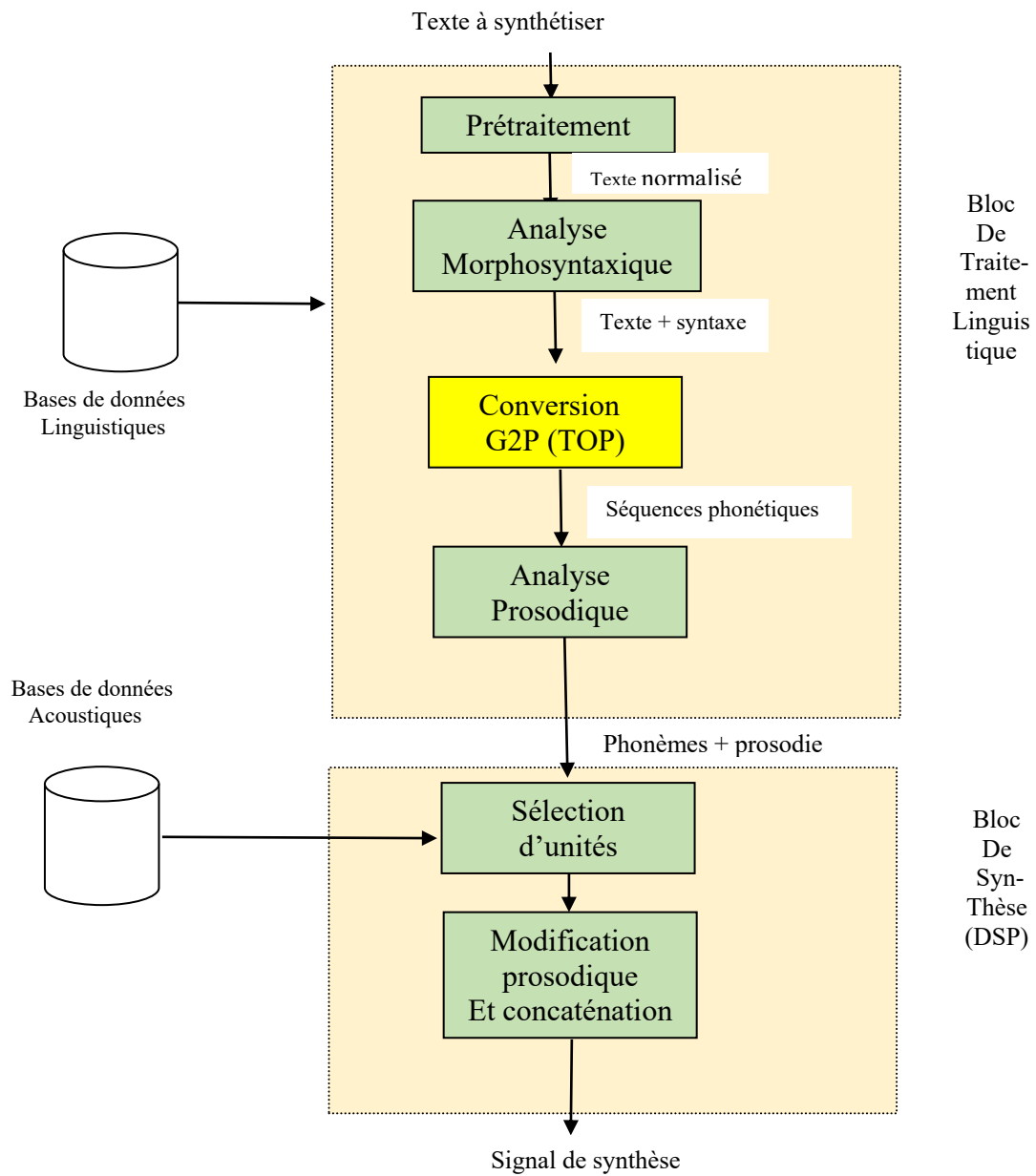


Fig. 1.10 : Architecture d'un système de synthèse de la parole à partir du texte (Synthèse par Concaténation) [46]

- *Phonétisation automatique (Conversion Graphème-Phonème)* : la Transcription Orthographique-Phonétique du texte ou phonétisation (qui fait l'objet de cette thèse) consiste à déterminer la séquence des phonèmes correspondant à la prononciation de ce texte. Cette conversion vise à créer la séquence phonétique associée à la séquence de graphèmes en sortie de l'analyse morphosyntaxique. Un graphème est la plus petite unité distinctive et significative de l'écriture. Dans l'écriture alphabétique, le graphème est communément appelé lettre. Pour la Conversion Graphème-Phonème, des règles phonétiques classiques de la langue choisie ainsi qu'une base contenant la transcription phonétique de certains mots irréguliers sont utilisées. La difficulté de l'opération réside dans le fait que la transformation n'est pas biunivoque : à une même séquence de caractères orthographique, ou graphème, peut correspondre une prononciation différente. Par exemple, pour le Français, le graphème x se prononce [s] dans « six », [ks] dans « axe » et [gz] dans « examen ». Réciproquement, un même phonème peut correspondre à plusieurs formes graphémiques. Par exemple le phonème [ε] s'écrit è dans « mère », ê dans fête, ei dans peine et ai dans « fait ». En Anglais, nous avons l'exemple du phonème [d] qui s'écrit th dans « the » et d dans « day ». Pour la langue arabe, ces points sont largement exposés aux chapitres suivants. Les difficultés précédentes reflètent des problèmes d'un niveau lexical ; un simple dictionnaire suffirait à fournir la transcription phonétique d'un mot. Cependant, prononcer une phrase, c'est-à-dire une séquence de mots, n'est pas qu'une simple transcription mot à mot. Des phénomènes d'ordre syntaxique et phonologique viennent modifier le contenu phonétique attendu d'une seule transcription lexicale. Nous pouvons citer, parmi les principaux phénomènes agissant sur la transcription au niveau phrase l'exemple de la coarticulation qui fait que la séquence des phonèmes (*chaîne phonémique*) est différente de la séquence des phones (*chaîne phonétique*). On peut tirer comme justification articulatoire à ces phénomènes un principe de moindre effort. Compte tenu d'une certaine inertie des articulateurs mis en jeu dans la phonation tous les phonèmes ne sont pas réalisés comme prévu, des variantes interviennent en fonction des stratégies d'élocution. Par exemple pour le Français, lorsque deux consonnes, qui diffèrent par leur trait de voisement, sont mises en contact, il en résulte une assimilation de ce trait : la séquence [sg] peut devenir [sk] (assimilation progressive) ou [zg] (assimilation régressive). Pour la langue arabe, les phénomènes qui altèrent la prononciation théorique des phonèmes (c'est-à-dire lorsqu'on passe à la réalisation sonore des phonèmes) sont formalisés au *chapitre 3*, nous pouvons en citer : *La gémination, l'emphatisation, l'arrêt glottal, l'assimilation partielle et l'assimilation totale*, etc.

Dans les différentes sections des *chapitres 2 et 3*, nous procédons à l'analyse des problèmes posés par la phonétisation automatique des textes de langue arabe et nous présentons une implémentation de ce traitement, basée sur une approche à base de règles, utilisant des dictionnaires d'exception (noms propres, mots étrangers, ...) et des règles de prononciation.

En fin de ce traitement, nous obtenons la transcription phonétique du texte en entrée du système de synthèse de parole, et qui sera injectée au bloc des traitements de synthèse du signal de parole. Les valeurs des paramètres prosodiques sont fournies quant à eux par le module de prédiction de la prosodie.

- *Prédiction de la prosodie* : le naturel d'un signal de parole synthétisé dépend en grande partie de la richesse prosodique qu'est capable de produire le système de synthèse de parole. Les paramètres prosodiques habituellement prédits sont le rythme (durées des segments phonétiques), l'intonation (les valeurs de la fréquence fondamentale F_0 des segments voisés) et l'intensité sonore (l'énergie). Ces paramètres apparaissent plus liés aux syllabes qu'aux segments phonétiques. C'est pourquoi ils sont souvent qualifiés de *suprasegmentaux*. La prédiction de la prosodie d'une phrase porte aussi sur la segmentation de cette dernière en groupes prosodiques, précisant ainsi la structure prosodique et les

positions des coupures (pauses) de cette phrase. Cette structure prosodique est relativement liée à la structure syntaxique de la phrase [47]. Les mots constituant une phrase ne sont pas tous équivalents. Certains mots peuvent recevoir un accent, pour le Français, toujours sur la dernière syllabe¹ (exemples : pâte, pas). Le module d'accentuation du système de synthèse peut utiliser plusieurs informations telles que la syllabation et la catégorie grammaticale des mots, pour placer ces accents. La prédiction de l'intonation utilise la structure des groupes prosodiques et les syllabes accentuées avec des modèles pour fixer l'allure de la courbe intonative [48, 49]. La prédiction du rythme, qui règle le débit de parole, peut utiliser des informations telles que l'identité du phonème, le contexte phonétique, la structure de la syllabe, la position du phonème dans la syllabe et de la syllabe dans le mot, etc, et des modèles qui peuvent être soit à base de règles définies manuellement [50], soit statistiques entraînés sur des corpus [51].

1.5.3.2 Traitements numériques : Synthèse par concaténation

La synthèse de parole par concaténation produit de la parole par la mise bout à bout de segments acoustiques (phones, diphones, syllabes, mots, etc.) puisés dans une base de données de parole préenregistrée par un locuteur et segmentée, en imposant au signal de parole synthétisé la prosodie prédite par le module de prédiction de la prosodie. La mise en œuvre d'une synthèse par concaténation nécessite des réponses à quelques questions relatives à la conception des Bases de Données de parole, à la sélection des unités dans ces bases, à leur modification prosodique et à leur concaténation. La *figure 1.11* présente le schéma général de la chaîne de ces traitements :

- *Bases de données de parole - Les unités de synthèse* : la construction d'une base de données de parole pour la synthèse par concaténation d'unités passe préalablement par faire un choix relatif à la nature de ces unités. Une caractéristique essentielle de ces unités dans la majeure partie des systèmes de synthèse de parole par concaténation, est leur définition basée sur l'identité phonétique. Cette dernière permet de faire le lien entre les données linguistiques issues du texte à phonétiser et les données acoustiques du signal de parole. Ainsi, on peut trouver des unités basées sur les phones (phone, n-phone, syllabe, mot) et d'autres basées sur les portions de phones (demi-phone, diphone, demi-syllabe, n-diphones).

Pour chaque langue, et pour chaque type d'unités acoustiques (phone, diphone, triphone, syllabe, mot et phrase), il faut avoir une estimation du nombre minimal de ces unités nécessaire pour couvrir l'espace acoustique de la dite langue.

Ainsi, la synthèse par concaténation de mots (ou de phrases) offre une qualité de synthèse supérieure, mais souffre d'un nombre d'unités acoustiques très important. C'est pour cette raison que la synthèse par concaténation de mots (ou de phrases) est réservée aux applications de la synthèse de parole à domaine limité [52]. La synthèse par concaténation de phones requiert, par contre, un minimum d'unités acoustiques dans la base de données (sous l'hypothèse que chaque phonème est représenté une seule fois dans cette base), mais au détriment d'une qualité de synthèse inintelligible. Ce fait est dû au faible degré de concaténabilité des phones (discontinuités importantes aux points de concaténation entre les phones).

Pour remédier au problème précédent, le phone, en tant qu'unité de synthèse, a été remplacé par le diphone [53].

Schématiquement, un diphone est une unité acoustique qui commence au milieu de la zone stationnaire (stabilité spectrale) d'un phone et se termine au milieu de la zone stationnaire du phone suivant. En somme, les diphones comportent les transitions entre phones. Historiquement, les diphones sont les premières unités acoustiques fidèles à la philosophie de la synthèse par concaténation. L'idée de cette synthèse consiste en effet à stocker les transitions entre phones au lieu de les modéliser (comme c'est le cas dans la synthèse par formants). Ainsi, en mettant bout à bout les parties semblables des phones, la synthèse par

¹- Comparativement à une syllabe non-accentuée, une syllabe accentuée est de durée plus longue, d'intensité plus forte et comporte d'importantes variations du F_0 .

concaténation de diphtonges élimine une grande partie des problèmes posés par les transitions entre phonèmes.

Contrairement au phonème, le diphtonge permet d'obtenir une parole intelligible, même si la base de données de parole ne contient qu'un seul représentant de chaque diphtongème de la langue [54]. Afin d'améliorer la qualité des systèmes de synthèse de parole par concaténation de diphtonges (des points de vue articulatoire et prosodique), les Bases de Données de parole de ces systèmes ont été enrichies en mettant dans ces bases plusieurs représentants de chaque diphtongème [55]. En conséquence, les systèmes de synthèse par concaténation se sont vus adjoindre un nouveau module dit de sélection d'unités. Malgré en effet l'amélioration des techniques de modification prosodique, ces dernières ne permettent toujours pas de compenser la faible variabilité acoustique d'une base de parole de petite taille (comme une base de diphtonges mono-représentés). Une modification trop importante des durées et des fréquences fondamentales de ces unités acoustiques génère une dégradation remarquable du signal de parole synthétisé.

Ce constat et un autre relatif au faible degré de concaténabilité de certaines classes de diphtonges, notamment celles couvrant les liquides et les semi-voyelles, ont conduit aussi à une autre évolution, qui consiste en l'utilisation de bases de parole de grande taille (une heure de parole et plus), avec la sélection des segments de parole les plus adéquats pour la synthèse, de manière à éviter autant que possible la modification prosodique du signal [56]. Cette synthèse est dite par concaténation d'unités de taille variable ou d'unités non-uniformes [57, 58, 59]. Le terme "non-uniforme" signifie que l'unité peut varier au niveau de sa longueur (diphtonge, phonème, demi-phonème, syllabe, mot, etc.) et au niveau de ses réalisations acoustiques (multi-représentation).

- *Préparation de la base de données des unités de synthèse* : la préparation d'une base de parole nécessite deux prises de décision primordiales, relatives au locuteur à enregistrer et au corpus de textes à vocaliser. La qualité de voix (timbre de la voix et manière d'élocution) du locuteur enregistré a une importance de premier ordre. C'est cette voix en effet qui sera restituée ensuite par le synthétiseur à partir des enregistrements. Habituellement, avant la phase d'enregistrement définitif du corpus de parole de la synthèse, on teste les qualités de voix de plusieurs locuteurs. Le locuteur final est généralement sélectionné sur la base de critères subjectifs (voix agréable, clarté de l'articulation et adéquation de la voix aux traitements de la synthèse) [60]. Le contenu du corpus de textes à vocaliser dépend de la stratégie de la synthèse. La synthèse par concaténation des unités d'un type donné nécessite la couverture de ces dernières dans le corpus de parole enregistré et par conséquent dans le corpus de textes à vocaliser. La synthèse par sélection d'unités nécessite en outre la présence de plusieurs variantes, notamment contextuelles (contexte phonétique), et si possible prosodiques (plusieurs intonations), de chaque unité. Actuellement, les corpus textuels sont constitués soit de logatomes, de mots isolés ou de phrases. Un logatome est un mot dénué de sens (séquence syllabique), qui permet de spécifier le contexte articulatoire (phonétique) de l'unité de synthèse. Chaque logatome contient généralement une seule unité de synthèse [54]. Les corpus textuels à base de logatomes sont notamment utilisés pour produire des bases de données de diphtonges. Cependant, la synthèse de parole à partir de ces dernières unités nécessite l'usage intensif de la concaténation, ce qui influe sur la qualité de la voix synthétique en donnant une impression de sur-articulation. Cette synthèse nécessite aussi l'usage intensif de modifications prosodiques quand ces unités sont mono-représentées. Comparativement aux logatomes, les mots isolés, et plus encore les phrases, offrent l'avantage de générer des corpus de parole prosodiquement proches de la parole naturelle. Les corpus textuels à base de mots isolés ou de phrases sont notamment utilisés pour produire des bases de données d'unités du type polyphone (ou n-diphtonge).

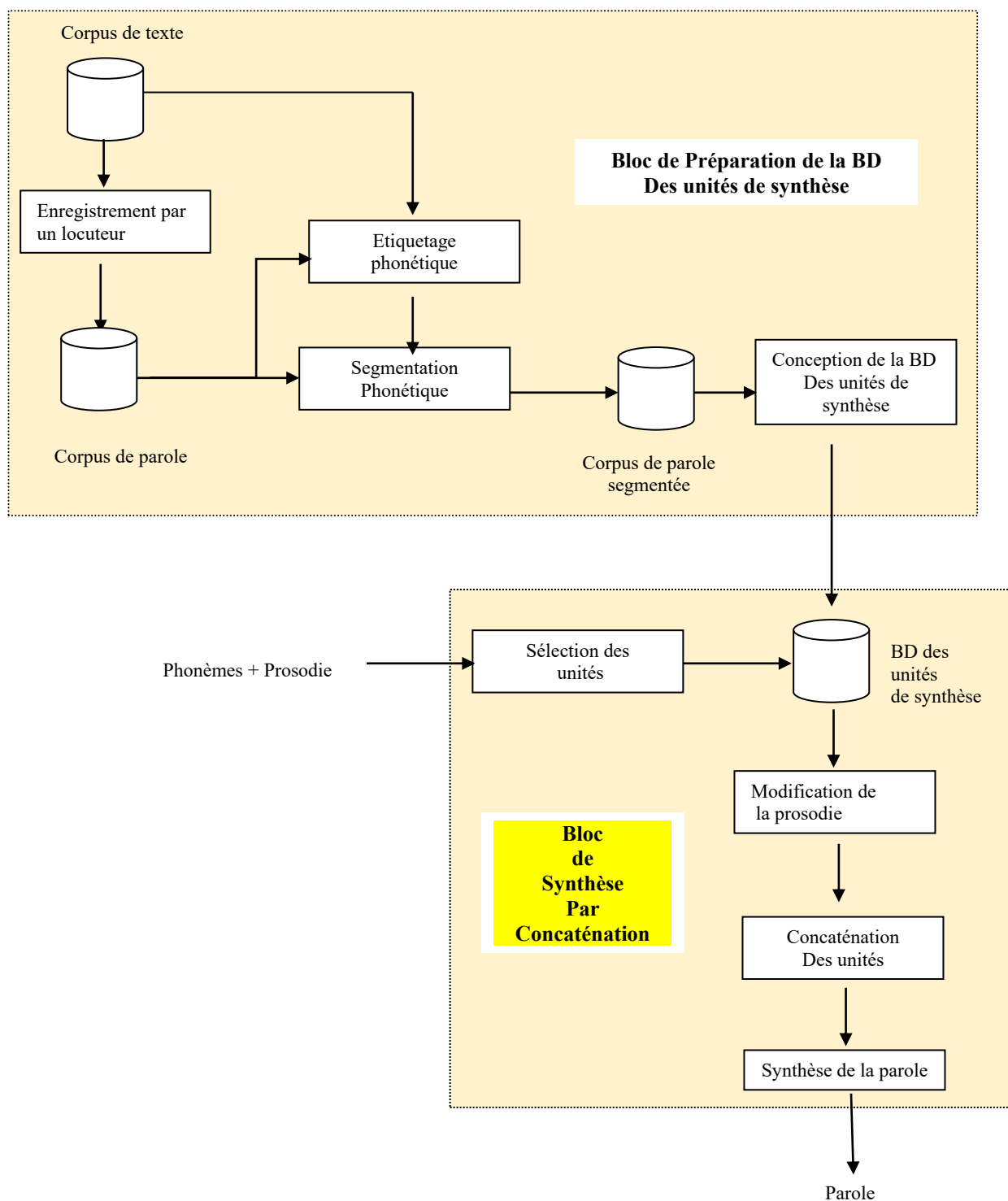


Fig. 1.11 : Représentation schématique des processus de préparation d'une Base de Données des unités de synthèse et d'un synthétiseur de parole par concaténation.

Comparativement aux diphtonges, ces unités permettent de réduire significativement le nombre de concaténations. Cependant, la multi-représentation de ces unités est nécessaire pour répondre aux contraintes du contexte phonétique des unités. Dans ce cas, la limite fondamentale d'un corpus textuel à base de mots isolés ou de phrases est sa taille, qui doit être la plus petite possible à cause des contraintes, par exemple, posées par les post-traitements que sont l'enregistrement et la segmentation. Pour cela, il existe des méthodes automatiques d'extraction, à partir de grands corpus de textes, d'un sous-ensemble de textes

représentatif des contraintes phonétiques et prosodiques des unités [61]. Le corpus de parole est généralement constitué par la lecture du corpus de textes. Ce mode d'élocution laisse au locuteur un degré de liberté quant à la prosodie employée (proche du naturel). L'enregistrement du corpus de parole peut prendre plusieurs jours, ou même plusieurs semaines. La voix du locuteur doit en effet être ménagée, car cette dernière peut être affectée si on impose au locuteur plus d'un certain nombre d'heures consécutives d'enregistrement. Une fois la phase d'enregistrement achevée, il s'agit ensuite de préparer, à partir du corpus de parole enregistrée, la base de données des unités de synthèse.

Pour cela, le corpus de parole enregistré est d'abord segmenté, par exemple, en phones, en syllabes ou en mots, suivant le type des unités du système de synthèse de parole.

La synthèse par concaténation de diphones par exemple, nécessite une segmentation de parole en phones [62, 63]. La segmentation en phones nécessite quant à elle l'étiquetage phonétique des énoncés. L'étiquetage et la segmentation phonétiques peuvent être exécutés manuellement pour de petits corpus de parole. Mais leur automatiser devient indispensable pour des corpus dépassant une certaine taille.

Une fois le corpus de parole segmenté, on procède à la conception de la base de données d'unités de synthèse. Ce traitement consiste à écarter tous les segments acoustiques non-exploitable par la synthèse, tels que les bruits et les valeurs acoustiques extrêmes, et à annoter symboliquement et acoustiquement (étiquettes phonétiques, durées, pitch, etc.) les segments acoustiques constituant la base des unités de synthèse, de manière à différencier ces unités. Cette annotation est ensuite utilisée par le bloc de sélection des unités [40].

○ *Synthèse du signal de parole*

Les informations phonétiques et prosodiques présentes en entrée du bloc de synthèse de parole (*fig. 1.10 et 1.11*) sont traitées par le module de sélection d'unités.

Dans le cas de bases de données d'unités multi-représentées, cette sélection est basée sur la minimisation d'un coût calculé sur toute la phrase à synthétiser. Pour cela, on met en œuvre des algorithmes basés sur la programmation dynamique, qui permettent de choisir les meilleurs segments acoustiques à concaténer, sachant que chaque cible spécifiée phonétiquement et prosodiquement en entrée peut avoir plusieurs unités candidates dans la base de parole. Ces algorithmes utilisent les paramètres symboliques et acoustiques des unités de la base de parole pour choisir parmi elles celles dont le contexte et la prosodie sont les plus proches de la chaîne phonétique et de la prosodie à synthétiser, et cela tout en minimisant un coût de concaténation qui représente la somme des discontinuités spectrales aux frontières des unités candidates à la concaténation [64, 57, 40, 59].

Une fois les unités choisies, ces dernières sont extraites de la base de données de parole et transmises éventuellement au module de modification prosodique qui ajuste ces unités suivant la consigne de prosodie générée par le module de prédiction situé dans le bloc des traitements linguistiques. Le principe fondamental de la synthèse par concaténation d'unités non-uniformes, sélectionnées dans de grands corpus de parole, est de réduire ou d'éliminer les sources de dégradation du signal synthétique telles que la modification prosodique [65].

Les techniques de modification prosodique sont de deux types [46] :

- Les techniques temporelles, initiées par l'algorithme TD-PSOLA (Time Domain-Pitch Synchronous Overlap and Add) qui a inspiré un autre algorithme appelé MBROLA (Multi Band Re-synthesis by Overlap and Add) [66]. Les techniques temporelles sont caractérisées par un coût calculatoire intéressant ;
- Les techniques fréquentielles ont commencé à donner des résultats satisfaisants avec l'apparition du modèle HNM (Harmonic plus Noise Model) [67]. Ces techniques présentent l'avantage de synthétiser le signal à partir d'une représentation paramétrique.

En sortie du module de modification prosodique, peuvent apparaître des discontinuités spectrales entre les unités acoustiques successives. Le traitement éventuel de ces discontinuités est confié au module de concaténation qui procède par un lissage spectral du signal de parole résultant de la concaténation de ces segments acoustiques successifs [67]. Une fois tous les traitements précédents effectués, le signal de parole constitué des segments concaténés est prêt à être synthétisé. Cette tâche de transformation physique est confiée au synthétiseur de parole à proprement parler.

1.6 Applications de la synthèse de la parole

Les applications des systèmes de synthèse sont nombreuses grâce à la demande croissante d'intégration de la parole dans plusieurs domaines. Le premier domaine d'application est les télécommunications. En effet, on cherche désormais à offrir le maximum de services à bas prix dans ce domaine. La synthèse permet de diminuer le coût des services accessibles par le téléphone comme les horaires de cinéma, le montant de sa facture téléphonique, annuaires, transactions boursières....

Les systèmes de synthèse peuvent aussi être un bon moyen de communication pour les handicapés. Ces derniers ont la possibilité de taper sur un clavier qui leur est adapté et de s'exprimer avec les autres personnes grâce à une voix de synthèse. On peut également imaginer l'utilité de ces systèmes pour des personnes malvoyantes qui peuvent avoir accès à des informations écrites grâce à une parole synthétisée. Cette dernière nécessite évidemment une reconnaissance de caractères. Une autre application intéressante est l'apprentissage des langues car la synthèse de la parole combinée à un outil d'apprentissage peut aider l'élève à apprendre une langue étrangère. La traduction automatique qui permet de traduire un mot d'une langue en une autre à voix haute est également un moyen d'apprentissage qui peut se faire par synthèse. Par rapport à ces services, la qualité de la parole synthétique doit être la meilleure possible.

Un autre domaine où la synthèse de la parole est employée est celui des jeux vidéo utilisant à la fois une interface jeu/joueur basée sur des technologies de reconnaissance vocale et des fonctionnalités basées sur la synthèse de la parole permettant aux personnages de jeux vidéo de dialoguer avec le joueur. D'autres domaines où la synthèse de la parole peut apporter une contribution significative sont les voitures et les maisons communicantes. Bientôt, les maisons et les voitures devraient pouvoir nous informer de certaines anomalies afin de diminuer le risque d'accidents. Pour tous ces domaines d'application l'intelligibilité est cruciale mais le naturel est d'autant plus nécessaire lorsque les systèmes de synthèse produisent des textes synthétisés longs.

1.7 Produits TTS commercialisés sur le marché

Cette section présente quelques-uns des systèmes de synthèse TTS commerciaux disponibles aujourd'hui. Plus de 28 systèmes de Synthèse TTS existant actuellement sur le marché, avec certains d'entre eux supportant la langue arabe. Une partie du contenu de cette section est basée sur des informations recueillies à partir des sites web officiels des développeurs, ou par contact direct avec les fabricants.

Les premiers systèmes de synthèse TTS commerciales étaient essentiellement basés sur le matériel et le processus de développement a été très long et coûteux (voir sections précédentes). Comme les ordinateurs se sont révélés être de plus en plus puissant, actuellement la plupart des synthétiseurs sont des systèmes basés sur des logiciels. Les systèmes à base de logiciels sont faciles à configurer et à mettre à jour, et le plus souvent sont beaucoup moins chers que les systèmes matériels. Toutefois, un dispositif matériel autonome peut encore être la meilleure solution quand un système portable est nécessaire.

1.7.1 PROJET MBROLA

Le projet MBROLA est l'un des principaux systèmes prenant en charge la langue arabe. Le projet MBROLA a été initié par le Laboratoire RTT de la Faculté Polytechnique de Mons, en Belgique. L'objectif principal de ce projet est d'avoir une synthèse de la parole pour autant de langues que possible. MBROLA est utilisé à des fins non-commerciales. Un autre objectif assigné par les concepteurs était d'aider la communauté scientifique à entamer des recherches en parole, en particulier dans la génération de la prosodie.

Le synthétiseur vocal MBROLA est basé sur la technique de concaténation de diphtongues. MBROLA produit des échantillons de parole sur 16 bits s'il est alimenté par une liste de phonèmes en entrée munis d'informations prosodiques. MBROLA utilise la méthode PSOLA (Pitch Synchronous Overlap Add) développée à l'origine par France Télécom (CNET). Il n'est pas, en fait, une méthode de synthèse en elle-même, mais permet la concaténation des échantillons de parole préenregistrés et offre un bon contrôle du pitch et de la durée.

Les bases de données de diphtongues dans MBROLA sont actuellement disponibles pour l'Anglais américain, Anglais britannique, Anglais Breton, le portugais brésilien, néerlandais, Français, allemand, roumain, espagnol, grec, gallois, langues indiennes, l'espagnol vénézuélien, hongrois, turc et pour l'AS. Certaines de ces langues existent avec les deux voix « mâle et femelle ».

1.7.2 GROUPE ACAPELA

Le Groupe Acapela rassemble toutes les technologies de la parole développées au cours des 20 dernières années. Des techniques de synthèse de la parole et de reconnaissance de la parole ont été améliorées et intégrées dans Acapela. Groupe Acapela a évolué à partir de la fusion stratégique de trois grandes entreprises européennes du domaine « des technologies vocales » : "Babel Technologies" créé au Mons, en Belgique, "Infovox" créé à Stockholm, en Suède et «Elan Speech" créé à Toulouse, en France. Acapela intègre actuellement trois technologies de pointe : le TTS par diphtongues, le TTS par sélection d'unités et la reconnaissance automatique de la parole. Acapela est actuellement disponible pour l'Anglais américain, l'Anglais britannique, la langue arabe, le néerlandais belge, le néerlandais, le Français, l'allemand, l'italien, le polonais, l'espagnol et pour le suédois.

1.7.3 ARABTALK

Le système de synthèse TTS ARABTALK a été développé à RDI (Research and Development International) en Egypte, et dédié entièrement à la langue arabe. ARABTALK est un système TTS concaténatif à base de corpus. Ce système utilise des modèles à base de réseaux de neurones artificiels (ANN) et des approches statistiques pour l'estimation de la prosodie, du pitch et d'autres paramètres nécessaires à la production de la parole de synthèse. En outre, il produit une synthèse en temps réel en exploitant un algorithme de sélection qui explore un large corpus de parole. ARABTALK exploite également des modèles de Markov cachés (HMM) pour l'alignement automatique dans le temps des transcriptions de nouvelles voix avec leurs phonèmes acoustiques correspondants. Le système est multi-utilisateur et adapté à l'exploitation client-serveur. ARABTALK couvre la plupart des phénomènes phonologiques de la langue arabe.

L'architecture globale et les caractéristiques générales du système TTS ARABTALK sont exposées à l'adresse : <http://www.rdieg.com/rdi/research/Arabtalk.asp>.

1.7.4 - Sakhr TTS

Le moteur TTS Sakhr convertit tout texte en arabe ou en Anglais en une voix humaine. Sakhr s'est efforcé au cours des 5 dernières années à créer un moteur TTS arabe qui peut correspondre à la qualité de la voix humaine. Cette technologie procure aux entreprises de télécommunication un avantage concurrentiel en leur permettant d'offrir à leurs clients les dernières informations statiques ou dynamiques - à tout moment et n'importe où - via les téléphones mobiles ou fixes.

Sakhr a intégré dans son système TTS un module de voyellisation automatique pour prendre en charge en entrée même les textes non voyellés et produire une parole de haute qualité en sortie. Le moteur TTS Sakhr est composé de trois parties. Le module linguistique convertit le texte d'entrée en une transcription phonétique. Le module phonétique calcule les paramètres de la parole et le module acoustique utilise ces paramètres pour générer des signaux de parole de synthèse.

1.7.5 Élan TTS

Élan TTS convertit tout texte en parole. Le logiciel est développé par la société Learnout & Hausple, un grand fournisseur mondial des logiciels Text-to-Speech et un développeur leader des solutions technologiques de pointe. Ce logiciel génère de la parole avec une bonne flexibilité et un naturel très convenable. Ce système fonctionne dans plusieurs langues avec des voix masculines et féminines en fonction de l'exigence de l'utilisateur. L'utilisateur peut choisir l'une des huit langues supportées par le logiciel, y compris l'AS. Ce logiciel est disponible sous les plateformes Windows NT, Linux, SCO, QNX, et Salaires.

1.7.6 FESTIVAL

Festival offre un environnement général intégré pour la construction de systèmes de synthèse de la parole et de tester leurs modules indépendamment.

Pour ce faire, FESTIVAL propose un certain nombre d'APIs: Allant du niveau Shell, passant par l'interpréteur de commandes, aux bibliothèques C ++, et Java, et terminant par une interface Emacs. Festival est multilingue (actuellement pour l'Anglais -britannique et américain-, et l'espagnol). D'autres groupes travaillent pour l'intégration de nouvelles langues y compris l'arabe standard. FESTIVAL offre également des outils et de la documentation pour construire de nouvelles voix complètes par le biais du projet Festvox de Carnegie Mellon (<http://festvox.org>).

Festival est un logiciel libre, il est distribué accompagné de quelques outils de la parole sous une licence de type X11 permettant aussi bien un usage commercial et non-commercial sans restriction.

1.8 Évaluation des systèmes de synthèse de la parole

L'évaluation de la synthèse de la parole est un domaine indispensable tant que la qualité de la parole produite sera inférieure à la qualité de la parole naturelle. En effet, l'évaluation de la qualité de la synthèse permet de déterminer certains défauts du système et d'y remédier.

Jusqu'à maintenant la grande majorité des tests utilisent les impressions subjectives de plusieurs sujets humains. Cette évaluation est appelée globale car elle juge la sortie du système de synthèse sans se préoccuper de son fonctionnement interne. L'autre type d'évaluation consiste à évaluer la qualité de chaque module du système synthèse. En effet, le système de synthèse est une chaîne de traitements, dont fait partie la phase de transcription Orthographique-phonétique, objet de notre thèse. Dans cette chaîne, le traitement qui fonctionne mal est celui qui va limiter les performances du système. Dans ce qui suit, nous présentons rapidement ces deux types d'évaluation.

1.8.1 Tests subjectifs ou évaluation globale

Globalement, les performances d'un système de synthèse se manifestent directement dans la parole générée. Cette dernière peut être jugée selon sa qualité, son intelligibilité, son naturel, son expressivité, la satisfaction des utilisateurs ou son adéquation pour l'application utilisée. Par exemple, pour des applications concernant les malvoyants, l'intelligibilité est plus importante que le naturel. Par contre, le naturel est le critère le plus important pour des applications multimédia ou dans les dialogues homme-machine.

Les méthodes utilisées pour évaluer la parole synthétique sont en fait des méthodes permettant de juger de la qualité d'un signal de parole quelconque. De plus, elles sont généralement subjectives car elles sont basées sur l'écoute et les avis d'un groupe de personnes. Il est donc clair qu'avant chaque test, certains éléments doivent être déterminés comme le type et le contenu des phrases à faire écouter, le scénario de test à suivre et le choix des auditeurs. Le choix du corpus de phrases à faire écouter dépend de la nature de l'évaluation et de l'application. Par exemple, pour

juger l'intelligibilité de la parole, certains tests utilisent des logatomes, des syllabes ou des phrases sémantiquement imprédictibles (Semantically Unpredictable Sentences ou SUS) formées des mots de la langue mais dépourvues de sens [68]. Les auditeurs sont ensuite priés d'écrire ce qu'ils ont entendu. Le corpus des phrases à faire écouter peut aussi contenir des cas de phonèmes, de mots ou groupe de mots qui sont problématiques dans la langue de l'application. Pour la langue arabe, par exemple, ce sont les mots commençant par un «alif lam» suivi par une lettre solaire etc... Le choix des auditeurs peut également modifier énormément les résultats de l'évaluation. Ces résultats risquent d'être erronés avec des auditeurs naïfs qui découvrent le test pour la première fois et seront plus représentatifs avec des auditeurs habitués au test. Ceci est dû au fait que l'oreille comprend de mieux en mieux au fur et à mesure des répétitions. Il peut être donc nécessaire de passer par une étape d'apprentissage afin que les sujets découvrent la qualité globale des phrases à tester.

Il existe plusieurs scénarios d'évaluation de la qualité de la parole et plus particulièrement de la parole synthétique. On trouve par exemple la méthode MOS (Mean Opinion Score) qui est incontestablement la plus utilisée et la plus simple [69]. Cette technique utilise une échelle de 5 niveaux qui qualifie la parole de très mauvaise (1) à excellente (5). L'auditeur n'a qu'à écouter le signal de parole puis mettre une note entre 1 à 5. Un autre scénario appelé CE (Categorical Estimation) consiste à évaluer la parole selon plusieurs critères : par exemple du moins ennuyeux au plus ennuyeux, du plus lent au plus rapide, du plus naturel au plus artificiel ou encore du plus facile ou plus difficile à comprendre. Ce scénario permet une analyse plus approfondie et donc aide à détecter les limitations et défauts du système. Il est possible aussi de comparer chaque couple de phrases du corpus et donc de noter une phrase par rapport à une autre.

Il existe aussi un scénario qui permet de juger de la qualité de la parole d'un système de synthèse par rapport à la parole issue d'un système de référence. Ce test est basé sur le jugement des dégradations (Degradation Category Rating ou DCR), où plusieurs personnes jugent sur une échelle de 1 à 5 le niveau de dégradation de la synthèse [70].

On peut également évaluer l'application en elle-même car la synthèse est souvent combinée à d'autres briques technologiques telles que des systèmes de reconnaissance ou de dialogue intelligent. Le moyen le plus pratique et efficace pour juger de la qualité de l'application est l'attente de retour d'utilisateurs réels. Ce n'est pas toujours possible car on préfère généralement tester le système avant son apparition sur le marché.

Il n'existe pas de test subjectif considéré par la communauté comme étant le meilleur. Il est en fait souhaitable d'utiliser plusieurs méthodes de test et de combiner les résultats d'évaluation. Il existe néanmoins des logiciels qui intègrent différents scénarios de test afin de faciliter la mise en œuvre des procédures d'évaluation de la qualité de la parole synthétique. Un de ces logiciels s'appelle SOAP (Speech Output Assessment Package) [71].

1.8.2 Tests objectifs ou évaluation locale

Comme tous les éléments d'un système de synthèse ont des rôles différents, ils influent tous obligatoirement sur la qualité de la parole produite. Avec une approche locale, les performances du système de synthèse sont déterminées à travers l'évaluation individuelle des performances de certains éléments clés du système comme la Conversion Graphème-Phonème, la prosodie, la sélection des unités et la concaténation. En effet, l'évaluation globale renseigne sur la qualité atteinte par les systèmes, en référence par exemple à la parole naturelle plus ou moins dégradée, ou à de la parole codée. Pour le concepteur du système, ou pour des applications spécifiques, il est nécessaire de tester également les composantes des systèmes séparément.

Un moyen d'évaluer la prosodie est, par exemple, de calculer une distance acoustique entre la prosodie générée par le système de sélection et la prosodie extraite du signal naturel. Les discontinuités au niveau des concaténations des diphtonges sont évaluées également par des distances.

1.9 Conclusion

Nous avons exposé dans ce chapitre l'état de l'art, l'historique, les fondements et les perspectives de la synthèse de la parole et des systèmes TTS en général, ainsi que leurs méthodes d'évaluation. Nous avons également donné un aperçu théorique rapide de la production de la parole, de la phonétique articulatoire, et d'autres concepts connexes. Cette présentation était indispensable pour une bonne compréhension du travail de cette thèse.

Chapitre 2:
Transcription Orthographique-Phonétique
en vue de l'Arabe Standard

2.1 Introduction

Traditionnellement appelée *Conversion Graphème-Phonème*, l'étape de Transcription Orthographique-Phonétique constitue le noyau minimal, indispensable à tout système de synthèse de la parole, aussi élémentaire soit-il. L'approche la plus convoitée pour cette étape repose sur l'utilisation d'un automate paramétré appliquant un ensemble de règles de réécriture, qui permettent d'associer un phonème (ou un groupe de phonèmes) à un caractère (ou un groupe de caractères) orthographique en prenant en compte le contexte gauche (caractère ou groupe de caractères précédant le segment à transcrire) et le contexte droit (caractère ou groupe de caractères suivant le segment à transcrire). Ces règles sont organisées de façon hiérarchique, des règles les plus particulières aux règles les plus générales [72]. Le nombre de règles nécessaires pour effectuer la Transcription Orthographique-Phonétique dépend de la langue que l'on considère. Certaines langues comme l'espagnol, le finlandais ou le swahili ont une correspondance plus ou moins directe entre l'écriture alphabétique et les systèmes sonores utilisés. Ces langues sont relativement faciles à transcrire en sons par de simples règles de transcription dépendantes de ces langues. D'autres langues, comme l'Anglais ou le Français, ne disposent que de régularités partielles entre leurs formes orthographiques et les systèmes sonores correspondants. La correspondance directe entre le système sonore et l'écriture orthographique n'est pas évidente pour l'Anglais et le Français, ainsi, pour de telles langues la Transcription Orthographique-Phonétique (TOP) basée sur des règles seules est une tâche fastidieuse [73].

Nous exposons dans ce chapitre, avec certains détails, les questions liées à la conversion du texte arabe en entités phonémiques et phonétiques pour être utilisées ensuite dans la synthèse de la parole ou dans d'autres applications. Les différentes approches de la TOP sont également étudiées en signalant les particularités de la langue arabe par rapport aux autres langues largement étudiées comme le Français et l'Anglais. Le développement de notre formalisme, son implémentation et l'évaluation de la qualité des résultats feront l'objet des chapitres suivants. La Phonétisation du texte -et les problèmes qui y sont associés- est une discipline qui a fait l'objet de recherches intensives pour d'autres langues comme l'Anglais et le Français en utilisant les différentes méthodes qui seront décrites dans la *section 2.3*, alors que les recherches pour l'AS sont relativement récentes. En outre, l'évaluation de la qualité de sortie de cette composante importante du TTS a été réalisée indépendamment des autres composants du système de la synthèse de la parole [74, 75]. Le présent chapitre prendra en charge, particulièrement, les points suivants :

- Les problèmes liés aux systèmes orthographiques des langues en général et de l'AS en particulier ;
- Les problèmes liés à la segmentation et au pré-traitement de l'AS, et les différentes approches de la transcription des langues en général et de l'AS en particulier.

2.2 Rôle de la TOP dans les systèmes de Synthèse de la parole

La TOP est une phase importante dans les systèmes TTS. La correspondance entre le système orthographique et la transcription sonore pour l'AS est à mi-chemin entre « simple » (comme pour l'Espagnol, le Finlandais ou le Swahili) et « complexe » (comme pour l'Anglais ou le Français). Les graphèmes de l'AS sont plus faciles à transcrire en sons à l'aide, par exemple, d'un ensemble de règles « lettre-son » renforcé par un lexique d'exceptions. Toutefois, les coarticulations dans la parole en langage naturel arabe peuvent inférer d'énormes variabilités contextuelles, comme c'est le cas de la pharyngalisation et de l'emphatisation et leurs influences sur les unités de synthèse retenues, et ceci nécessite une attention particulière [76]. Ainsi, le problème de la conversion « Lettre-son » pour l'AS n'est pas seulement de convertir les graphèmes en entités linguistiques abstraites de base comme les phonèmes, mais plutôt la conversion en phones (allophones) représentant les sons réels de la langue [77]. Ceci exige qu'un autre ensemble de règles doive être déployé pour couvrir les importantes variations phonétiques de l'AS en parole continue [78].

La Phonétisation des textes est importante pour la synthèse et la reconnaissance vocale et pour d'autres applications de Traitement Automatique de la Parole (TAP). En synthèse de la parole, des règles servent à établir la correspondance entre l'orthographe et les sons (phonèmes et phones ou allophones). Les sons peuvent ainsi être utilisés seuls ou convertis en syllabes, qui peuvent être en plus divisés en sous-groupes appelés clusters (nous avons un cluster lorsqu'il y a une règle phonologique permettant de scinder deux sons adjacents en une unité indépendante), et utilisés pour la synthèse de la langue. En reconnaissance de la parole, les règles « lettre-son » sont utilisées pour générer des variantes de prononciation afin d'améliorer la qualité de la reconnaissance et générer les prononciations de nouveaux mots add-on, qui n'existaient pas dans le vocabulaire d'origine du système de reconnaissance vocale. Et comme application dans le domaine TALN en général, la Conversion Graphème-Phonème peut être utilisée à des fins pédagogiques telles que la correction de fautes d'orthographe. Une solution raisonnable au problème de la conversion « lettre-son » est de disposer d'un système à trois composantes :

- le module de segmentation du texte et de pré-traitement pour formater le texte d'entrée en éléments lexicaux bien structurés et convertir les abréviations, les symboles, les acronymes et les numéros en séquences de mots ;
- le module de transcription graphème-phonème proprement dit, qui transcrit les graphèmes en séquences de phonèmes ;
- une composante de transcription phonétique qui transcrit les phonèmes en phones (allophone = des sons réels). En effet, un Phone ou allophone est la réalisation sonore d'un phonème. Cette réalisation peut-être en plusieurs variantes, d'où la nécessité de disposer de règles supplémentaires appelées règles phonémiques ou phonologiques.

La sortie de ces deux derniers modules est une transcription phonétique du texte d'entrée. La Segmentation du texte et son pré-traitement sont une interface pour tout système de phonétisation de texte et nous les exposerons dans la *section 2.4.6*, après avoir abordé les problèmes liés au système d'écriture de la langue arabe standard. Les composantes de transcription phonologique et phonétique seront détaillées dans le chapitre suivant.

2.3 Méthodes de transcription « Lettre-son »

Il existe trois méthodes utilisées pour la transcription « lettre-son » de la plupart des langues. Ces méthodes sont les suivantes :

- des méthodes à base de dictionnaires qui reposent sur le stockage du maximum d'informations phonologiques des morphèmes dans un lexique ;
- méthodes à base de règles où des informations linguistiques et phonétiques pertinentes sont utilisées pour élaborer un ensemble de règles « lettre-son » renforcé par un lexique des exceptions nécessaire lorsque ces règles ne peuvent plus s'appliquer ;
- les méthodes relativement nouvelles appelées *data-driven methods*, Méthodes guidées par les données.

2.3.1 Transcription à base de Dictionnaire

La transcription lettre-son à base de dictionnaire repose sur le stockage- dans un lexique- du grand nombre possible de connaissances phonologiques (y compris la prononciation des morphèmes). Les morphèmes sont utilisés à la place des mots afin de réduire la taille du lexique. La prononciation des mots d'entrée est générée à partir des morphèmes (stockées) grâce à des règles morphologiques complexes incluant des règles flexionnelles et dérivationnelles et en composition avec des règles morpho-phonémiques, qui décrivent comment la transcription phonétique des constituants morphémiques varie quand ils sont combinés en mots. La méthode à base de dictionnaire, pour l'Anglais, était élaborée pour la première fois par *Coker* en 1985 [79], puis par *Allen et al.* en 1987 [3], ensuite perfectionnée par *Coker et coll.* en 1990 [80] et finalement par *Levinson et al.* en 1993 [81]. Pour le Français, un dictionnaire de taille importante a

été créé par *Laporte* en 1988 et utilisé pour la conversion lettre-son [82]. Des solutions pratiques à base de dictionnaires aux problèmes de la transcription « lettre- son » pour l'Anglais ont été intégrées dans le système *MITALK TTS* [3] et dans le système TTS de *AT& T* [81].

Les entrées lexicales dans le dictionnaire peuvent contenir des informations graphémiques, phonétiques, syntaxiques et sémantiques. Un dictionnaire complet nécessite une énorme mémoire informatique et un effort fastidieux pour son élaboration. Son principal avantage est la possibilité de son exploitation dans d'autres applications telles que l'étiquetage et l'analyse grammaticale des phrases nécessaires à l'amélioration de l'intonation et du naturel des synthétiseurs vocaux. Il peut avoir également des applications dans le domaine de la traduction automatique ou la reconnaissance automatique de la parole.

L'AS est une langue fortement flexionnelle de sorte qu'une racine en langue arabe peut avoir plusieurs formes fléchies (une moyenne de dix à douze formes et quatorze formes exceptionnellement). La racine "درَس" (l'acte d'étudier), par exemple, possède 11 formes : quatre pour les verbes et sept formes nominales. Un dictionnaire orthographique complet de la langue arabe doit tenir compte de toutes ces formes, ce qui augmente considérablement le nombre de ses entrées. Par exemple, le dictionnaire lexical arabe exhaustif d'*Al Qamous Almuhiy*, compte plus de 250 000 entrées [83]. Heureusement, toutes ces formes dérivées respectent de strictes règles orthographiques. Ajoutant à cela la relative simplicité du système orthographique de la langue arabe qui jouit d'une bonne correspondance avec la prononciation réelle des mots, ce qui fait du système de transcription à base de règles, soutenue par un dictionnaire de mots exceptionnels, une solution viable au problème de conversion lettre-son pour la langue arabe standard.

2.3.2 Transcription à base de règles

Les systèmes de transcription à base de règles utilisent un ensemble complet de règles graphème-phonème et un dictionnaire des exceptions (les mots qui constituent leurs propres règles). Les catégories de mots exceptionnels de l'AS que nous avons retenues dans notre approche seront décrites au cours des sections qui suivent, ainsi qu'un post-processeur phonétique qui transcrit le texte en sons réels. Depuis l'émergence des méthodes à base de règles, des efforts progressifs considérables ont été déployés pour concevoir des ensembles de règles et d'exceptions à large couverture. Pour l'Anglais et le Français ceci est réalisé par *Ainsworth* [84], *McIlroy* [85], *Harrison et al.* [86], *Hertz* [87], *Hunnicut* [88], *Belrhali et al.* [89] et *Divay et Vitale* [90]. Les systèmes à base de règles les plus élaborées sont ceux à base de connaissances d'experts, car ils utilisent les connaissances linguistiques et phonétiques des experts de la langue. Les différents types de formalismes de règles sont liés aux aspects suivants : les différences en nombres de règles, l'inventaire des phonèmes, les types et les formats des règles, la façon dans laquelle les règles sont analysées, la taille du dictionnaire des exceptions, l'algorithme utilisé pour l'analyse de ce dictionnaire, etc... Les méthodes à base de règles sont un langage spécifique et sont largement utilisés dans la synthèse de la parole.

Les méthodes à base de dictionnaire et celles à base de règles ne sont pas mutuellement exclusives. La transcription à base de règles utilise également un dictionnaire des exceptions. La différence entre les deux méthodes réside dans l'importance relative à l'exploitation de l'intelligence phonologique. Dans le but d'une correspondance régulière entre l'orthographe et la prononciation en MSA, une intelligence phonologique maximale doit être intégrée dans ces règles, et le problème de la phonétisation de la langue arabe peut être résolu par l'élaboration d'un ensemble complet de règles phonologiques et phonétiques, assisté par un dictionnaire des exceptions. D'autre part, les langues comme l'Anglais ou le Français ont des systèmes orthographiques complexes, ce qui signifie que l'élaboration de règles phonologiques génériques complètes contextuelles est presque impossible. Ceci nous oblige à intégrer dans le dictionnaire le maximum de connaissances linguistiques plutôt que de simples règles phonétiques ou phonologiques, c'est à dire, la taille de ce dictionnaire pour les langues comme l'Anglais ou le Français est assez conséquente. Une autre différence entre le dictionnaire d'exceptions utilisé dans les méthodes à base de règles et celui utilisé dans les systèmes de transcription à base de dictionnaire est dans le type d'informations stockées. Comme nous l'avons souligné plus haut, les informations contenues dans le dictionnaire d'une méthode à base de dictionnaire sont complexes

et peuvent comporter des morphèmes et leurs prononciations et éventuellement d'autres informations syntaxiques et sémantiques. Dans le dictionnaire des exceptions utilisé dans la méthode à base de règles, les informations stockées sont uniquement les mots exceptionnels et leurs prononciations. L'approche mixte de règles et de dictionnaire a été exploitée à fond pour l'Anglais et le Français et peut être utilisée pour la langue arabe et où, dans les deux cas, le dictionnaire doit être analysé avant l'exécution des règles de conversion « lettre-son ».

Compte tenu de la nature du système d'écriture de l'AS et la relation régulière entre l'orthographe et la prononciation où la correspondance entre les graphèmes et les phonèmes est quasi-biunivoque, il semble qu'il est bien judicieux d'adopter la transcription à base de règles, car il est possible de développer un ensemble généralisé de règles « lettre-son » qui couvrent la majorité de l'orthographe arabe. Le dictionnaire des mots exceptionnels portera sur les exceptions aux règles générales. En effet, Ce dictionnaire est consulté, au cours du traitement, avant toute recherche de règles : Si le mot à phonétiser est présent, il est transcrit directement et l'étiquette *EXCEPTION* lui est opposée. Ce dictionnaire est codé sous la forme d'un arbre en partie commune sur les graphies afin de garantir un temps constant d'accès aux éléments quelle que soit la taille du lexique. Dans l'élaboration des règles de transcription orthographique-phonétique, il est supposé que les mots sont correctement orthographiés, c'est à dire, le texte à vocaliser en langue arabe est soit correctement disponible ou bien son *accommodation* est facilement générée par un module NLP séparé [91, 92, 93]. Puisant dans la littérature abondante sur la langue arabe et son orthographe, il est possible de compiler un ensemble de règles précises pour transcrire le texte arabe en phonèmes et en phones [94, 95]. L'Anglais est une langue germanique qui a puisé énormément des langues latines. Contrairement à l'arabe, l'Anglais a un système orthographique complexe et nécessite des centaines de règles « lettre-son » pour traduire correctement environ 26% des mots en situation de texte illimité comme Damper l'a montré dans [96]. Pour l'Anglais, beaucoup de mots fréquemment utilisés (comme *the*, *of*) violent souvent les règles de prononciation de base et doit être traités par des listes d'exceptions. Ces tourments n'existent pas en langue arabe. En mettant de côté les quelques exceptions aux règles, la langue arabe peut être considérée comme une langue phonétique ayant un système orthographique régulier (elle est parmi les rares langues orthographiquement correctes). À cet égard, la langue arabe peut être modélisée phonétiquement avec des règles spécifiques développées manuellement à l'aide de notre compétence linguistique et phonétique. Nous présenterons dans le chapitre suivant les règles principales, avec leurs sous-règles, pour la Conversion Graphème-Phonème et 22 règles supplémentaires de transcription phonème-phone. La comparaison est dérisoire avec le Français qui nécessite 500+4000 règles !!! [97]. Si la liste des exceptions est exhaustive et les règles « lettre-son » sont complètes et correctes, la transcription de texte arabe serait alors précise. Pour la langue arabe standard, l'association des règles « lettre-son » appropriées avec un dictionnaire de mots exceptionnels (et d'autres lexiques annexes d'appui), est généralement suffisant pour atteindre une haute précision de transcription « lettre-son ». Un grand nombre d'erreurs de transcription phonétique est dues aux noms propres et acronymes empruntés des langues étrangères. Techniquement, la langue arabe est influencé par les deux langues : l'Anglais et le Français, et en général, les mots étrangers empruntés d'autres langues doivent être transcrits dans le système orthographique d'origine et obéissent à leurs propres règles de transcription « lettre-son ».

2.3.3 Transcription Data-driven

Il existe trois approches relativement nouvelles à base de données (Data-Driven approaches) : la prononciation par analogie (PbA), des méthodes statistiques basées sur la théorie stochastique et le principe du plus proche voisin, et les méthodes connexionnistes basées sur les réseaux de neurones [98]. L'approche PbA gagne actuellement en popularité. Elle exploite les connaissances phonologiques contenues implicitement dans un dictionnaire de mots et leurs prononciations correspondantes. Nous pouvons citer comme exemples de l'approche PbA, l'œuvre de Dedina et Nusbaum (1991) [99], de Yvon [100, 101], de Damper et Eastmond [102], Bagshaw (1998) [103] et Marchand et Damper (2000) [104]. Un état d'art de ces méthodes « Data-Driven » est présenté de manière exhaustive dans l'ouvrage de Damper [105]. L'idée sous-jacente dans l'approche PbA

est la détermination de la prononciation d'un nouveau mot à partir des parties similaires des mots connus et leurs prononciations correspondantes. Ainsi la prononciation d'un nouveau mot inconnu est assemblée après une recherche de correspondance entre les sous-chaînes du nouveau mot d'entrée et des chaînes des mots lexicaux connus dans le dictionnaire.

Une prononciation partielle est supposée pour chaque sous-chaîne correspondante de la connaissance phonologique, et les prononciations partielles sont concaténées. L'approche PbA nécessite un dictionnaire dans lequel les formes orthographiques des sous-chaînes de mots du dictionnaire sont alignées avec les prononciations correspondantes de sorte que ces sous-chaînes de correspondance sont facilement identifiables.

Les méthodes statistiques comme celle de la transduction stochastique [106], ou celle du « plus proche voisin » [107], sont des approches Data-Driven avec entraînement. Dans le procédé IB1-IG exposé dans [107], par exemple, un matériau d'entraînement comme la correspondance Graphème-Phonème est utilisé pour générer une classification de phonèmes avec une certaine probabilité. Le procédé IB1-IG est fondé sur une méthode de classification automatique. Ce procédé est une technique « *Lazy Learning* », apprentissage paresseux, de classification. On dit qu'une technique est « *Lazy learning* » quand la généralisation des données du système est retardée jusqu'à ce qu'une requête soit faite au système, par opposition à « *Eager learning* », apprentissage pressé, où le système essaie de généraliser les données d'entraînement avant de recevoir des requêtes. Ce dernier est donc un apprentissage « hors ligne » dans lequel les requêtes post-formation adressées au système n'ont aucun effet sur le système lui-même. Par contre, « *Lazy learning* » ne construit pas de modèle à partir de données d'entraînement, mais pour prédire la classe d'une nouvelle instance, il compare aux instances stockées à l'aide d'une métrique de similarité, et la nouvelle classe d'instance est déterminée en fonction des classes des instances d'entraînement les plus similaires. Au niveau algorithmique, les algorithmes de « l'apprentissage paresseux » sont des versions de classificateurs « k-plus proche voisin » (K-NN), où le poids de similarité d'une instance est automatiquement calculé en tant que Gain d'Information (IG) de cette instance. Les Réseaux de Neurones à base de Perceptrons Multicouches (MLP) avec l'algorithme de rétro-propagation pour l'apprentissage ont également été utilisés dans la transcription des textes comme ceux développés par Sejnowski et Rosenberg [108], et Matsumoto et Yamaguchi [109]. Les solutions basées sur les MLP sont indépendantes de la langue, mais elles présentent quelques inconvénients tels que le traitement des groupes (clusters) de graphèmes et les caractéristiques syntaxiques [4]. Cherifi et Guerti dans [110] ont développé une méthode basée sur les transducteurs à états finis pondérés. Les mêmes auteurs, dans [78] ont exploité une approche probabiliste avancée basée sur les données, à savoir les champs aléatoires conditionnels (CRF). Ces mêmes auteurs, ont proposé dans [77] une approche pour la Conversion Graphème-Phonème en AS, basée sur une autre technique « Data-Driven » : le modèle multi-gramme conjoint (JMM). Dans ce genre d'approches, nous n'avons pas besoin d'explicitement toutes les anomalies de la correspondance Graphème-Phonème, mais toutes les connaissances Orthographico-Phonétiques seront incluses implicitement dans l'étape d'entraînement du modèle.

Il existe d'autres approches hybrides (à base de règles et Data-Driven) comme les travaux de Meng dans [111] qui peuvent être utilisés pour la Conversion Lettre-Son et également la conversion Son-Lettre, ce qui signifie la possibilité de leur exploitation à la fois pour la synthèse comme pour la reconnaissance vocale.

Les systèmes à base de dictionnaires sont des solutions complexes à des langues à systèmes orthographiques complexes et irréguliers comme pour l'Anglais et le Français. L'avantage principal des méthodes à base de dictionnaire et celles à base de données est leur indépendance vis-à-vis de la langue. Toutefois, pour introduire de telles méthodes à une nouvelle langue comme la langue arabe standard, des recherches considérables et des efforts manuels sont nécessaires pour créer le lexique ; et pour l'introduction des méthodes comme PbA, un effort supplémentaire est nécessaire pour aligner le texte aux prononciations. Dans cette perspective, les avantages de l'utilisation de la transcription à base de dictionnaire ou celle de la PbA pour les langues à système de prononciation simple comme la langue arabe, doivent être soigneusement pesés contre ceux de la transcription à base de règles.

2.4 Problèmes de la Conversion Graphème-Phonème

Au niveau phonémique, la transcription orthographique arabe se caractérise par certains problèmes tels que : la normalisation de texte, des problèmes morphophonémiques, élision, les noms propres, les nouveaux mots et les mots étrangers à la langue et les irrégularités orthographiques. Il est à noter que ces problèmes sont communs à d'autres langues comme l'Anglais ou le Français [90]. Les problèmes d'orthographe par exemple, tels que les différentes réalisations phonétiques de certaines séquences de graphèmes sont rencontrés en Anglais et en Français. Les problèmes morpho-phonémiques tels que la dépendance de la réalisation phonétique d'un mot par rapport au mot précédent et au mot suivant existent également en Anglais et en Français. L'élision est spécifique au Français, mais dans une mesure limitée, elle existe aussi en langue arabe. Les néologismes et la prononciation des noms propres sont des problèmes presque universels. Au niveau phonétique, l'ensemble des règles de transcription phonèmes-phones est spécifique à la langue arabe. Un élément important de la transcription phonème-phones sont les règles de pharyngalisation de la langue arabe (*section 3.2.2* du chap. 3). L'entrée du module de pharyngalisation ainsi que celui de la génération des diphtongues, exige que la séquence phonémique, générée par l'application des règles lettre-phonème, soit décomposée en syllabes. C'est le processus de syllabation des phonèmes de la langue arabe qui sera traité à la *section 2.4.5*.

2.4.1. Système d'écriture de l'AS

La langue est un système conventionnel constitué de symboles et de signes exploités par les Hommes pour communiquer entre eux, et pour exprimer leurs idées. Nous pouvons également définir une langue comme étant un ensemble de sons produits par l'appareil phonatoire humain, et perçus par l'oreille, engendrant ainsi certaines connotations idiomatiques reconnues dans la communauté désignée. La langue en tant qu'ensemble de sons - une parole - préexistait bien avant la langue en tant que système orthographique - l'écriture-. En effet, l'écriture n'est apparue que depuis 6000 ans dans deux contrées voisines, la Mésopotamie et l'Égypte, de manière presque simultanée mais différenciée. Si les hiéroglyphes égyptiens et les pictogrammes sumériens sont tous les deux formés de petites images, celles-ci sont totalement propres à leur région. Continuant à se répandre dans le monde, l'écriture va utiliser de nouvelles règles : c'est l'invention de l'alphabet.

L'alphabet se compose d'un ensemble conventionnel de signes écrits dont chacun correspond à un seul son parlé ; tous ces signes, dont le nombre est limité, sont susceptibles d'être disposés selon des combinaisons interchangeable de façon à former des diverses syllabes et les différents mots. L'écriture semble avoir été inventée vers 3400 avant l'ère actuelle à Ougarit, un port de commerce alors très actif, où on a découvert en 1928 une série de tablettes écrites à l'aide de 30 signes seulement, d'aspect cunéiformes, utilisés pour noter des sons et non plus des idées.

Il en a résulté de ce retard dans la transcription graphémique de la parole une certaine confusion entre les deux concepts : parole et texte, c'est-à-dire entre ce qui est prononcé et ce qui est transcrits en termes de symboles représentant ces sons. Cette confusion n'est apparemment pas récente, les anciens linguistes arabes en étaient une victime : certains d'entre eux ne distinguaient pas entre le son et la lettre, ils utilisaient parfois le terme « lettre » pour désigner « le son prononcé » - le phonème - et parfois pour désigner son image visuelle graphémique -l'écrit-. Nous nous contentons ici de donner un seule exemple, celui de l'étude d'Ibn Sina (838 AH), « *les causes de production des lettres* », (أسباب حدوث الحروف) dans laquelle la confusion entre les deux concepts est visible. Ajoutez à cela que certaines personnes présument que les mots visuelles, c'est-à-dire ces symboles dessinés, reflètent fidèlement leur langue parlée, mais la réalité en est autre chose. En effet, la pratique « quotidienne » de l'écriture alphabétique conduit à accepter comme définition tout à fait évidente de l'écriture : *système de représentation des productions de la langue parlée*, ou encore : *transcription de la chaîne parlée*. Cependant, certaines réalités pourraient jeter un doute sur cette définition et cette conviction. Il suffit d'essayer de transcrire scrupuleusement, exactement, un échange oral pour se rendre compte que celui-ci ne peut pas être

écrit « tel quel », qu'il doit être transformé plus ou moins profondément pour être simplement intelligible à la lecture (ou inversement d'ailleurs). Néanmoins, Il faut faire la distinction entre les deux termes "écriture" et "orthographe". En effet, l'écriture étant antérieure à l'orthographe, elle est la tentative d'un usager d'une langue à calligraphier (représenter graphiquement) la langue, indépendamment de toute règle conventionnelle, tandis que "l'orthographe" est un système de normalisation et de description de l'écriture, où certaines formes sont validées, alors que d'autres, en sont exclues. La relation « écriture-orthographe » est analogue à la relation « langue-grammaire ».

L'alphabet arabe a des origines sémitiques dérivé du système d'écriture araméen, et compte parmi les plus anciens alphabets du monde [112,113]. Le système d'écriture arabe renferme un alphabet régulier pour les consonnes, des signes diacritiques pour les voyelles et autres signes utilisés. Le système orthographique qui régit l'écriture de la langue arabe, et dans certains cas, ne repose sur aucune base formelle, seule la prévalence de l'usage qui l'emporte. Il y a six voyelles pour la langue arabe : les courtes en nombre de trois représentées phonétiquement par /a/, /u/, /i/ et trois longues voyelles correspondantes, représentées par /a:/, /u:/ et /i:/.

La langue arabe possède 28 consonnes. En utilisant le « Speech Assessment Methods Phonetic Alphabet » (SAMPA) pour la langue arabe [114], les sons consonantiques peuvent être représentés par : ʔ, b, t, T, ʒ, x, X, d, D, r, z, s, S, s', d', t', D', ʔ', G, f, q, k, l, m, n, h, w, et j. Développé dans le cadre des projets européens Esprit SAM, SAMPA constitue un standard bien établi, faisant autorité dans les milieux du traitement automatique des langues naturelles et de la parole [114]. Rappelons que SAMPA est un système permettant de coder les symboles API standard à partir de symboles ASCII. Néanmoins, concernant la désignation des termes relatifs au domaine phonéto-orthographique de l'AS, nous adoptons parfois la translittération en caractères latins, tels que : alif, idgham, damma, fatha, etc., mais lorsqu'il s'agit de la transcription formelle Graphème-Phonème, nous nous restreignons aux exigences de la codification adoptée SAMPA, et nous aurons alors : [ʔalif], [idGa:m], [d'amma], [fatxa], etc. Le système d'écriture de l'AS se compose de :

- trois voyelles diacritiques qui apparaissent au-dessus des graphèmes représentant les consonnes, ou au-dessus de Chedda (voir le troisième point). Les voyelles diacritiques indiquent que la consonne sur laquelle apparaissent ces signes, est vocalisée, comme par exemple dans le mot « جَيِّدٌ » [ʒajjid] (bon) qui a une chedda et une kasra (marque de la voyelle diacritique ِ) toutes les deux présentes sur la semi-voyelle [j] (ي) ;
- vingt-huit graphèmes représentant les sons des consonnes ;
- la chedda « ُ » ou signe de gémiation. Ce signe apparaît normalement sur une consonne pour indiquer qu'elle est géminée, c'est-à-dire que le son correspondant est répété (voir l'exemple du premier point). En effet, l'école traditionaliste des grammairiens arabes considère que la gémiation est un dédoublement de la consonne, c'est à dire prononcer une consonne d'une manière appuyée [115] ;
- trois symboles du « [tanwi:n] » : [tanwi:n fatxa], [tanwi:n kasra] et [tanwi:n d'amma]. Ils apparaissent sur le [ʔalif] ou toute consonne pour indiquer certaines séquences de phonèmes, comme par exemple dans les mots suivants : " كَبِيرًا " [kabi:rin] (énorme), " صَغِيرٌ " [s'aGi:rin] (petit) et " جَمِيلٌ " [ʒami:lun] (beau) ont les symboles du [tanwi:n] ;
- quelques symboles de ligatures comme [ʔalifla:m], [la:mʔalif], etc., comme par exemple dans les mots " الْبَيْتُ " [ʔalbajti] (la maison) et " لَانٌ " [laʔan] (car) où les symboles de ligature sont [ʔalifla:m] et [la:mʔalif], respectivement. Dans [116], Yannis Haralambous définit trois types de ligatures : linguistiques, esthétiques et contextuelles.

Les symboles des voyelles sont :

- fatha, un signe diacritique, qui ressemble à un trait d'union, "-". Il apparaît au-dessus d'une consonne pour indiquer qu'elle est vocalisée par le son de la voyelle courte [a] ;
- kasra, un signe diacritique, qui ressemble à un trait d'union et qui apparaît sous une consonne pour indiquer que cette consonne est vocalisée par le son de la voyelle courte arabe [i] ;
- damma, un signe diacritique, qui ressemble à une virgule et qui apparaît en haut d'une consonne pour indiquer que cette consonne est vocalisée par le son de la voyelle courte [u].

Les consonnes, quand elles apparaissent seules, correspondent aux sons non vocalisés de la langue arabe. Dans la pratique du langage, les consonnes sont généralement vocalisées par la présence d'une voyelle ou d'un signe diacritique [tanwi:n] au-dessus de leurs graphèmes. La présence du signe de gémination " chedda " au-dessus des signes diacritiques des consonnes indique que ces sons doivent être répétés. Les graphèmes arabes, correspondant aux sons des consonnes, sont respectivement :

(ي، و، هـ، ن، ل، ك، ق، ف، غ، ع، ظ، ط، ض، ص، ش، س، ز، ر، ذ، د، خ، ح، ج، ث، ت، ب، ء)

Mais pour les besoins de la transcription les 28 consonnes arabes ont été divisées en deux groupes (voir *Tableau 2.4*) :

- 14 consonnes solaires qui assimilent le « َ » de l'article ;
- 14 consonnes lunaires qui n'assimilent pas le « َ » de l'article.

Le *Tableau 2.1* montre l'alphabet phonétique des sons de l'AS en utilisant SAMPA pour l'alphabet arabe (la version étendue de SAMPA est exposée en annexe 2). Les *Tableaux 2.2* et *2.3* représentent les articulations des consonnes arabes et des sons de voyelles. Le système d'écriture arabe utilise également quelques symboles spéciaux et certains signes de ponctuation. Une caractéristique exclusive de l'écriture en langue arabe est que les graphèmes sont connectés, même en caractères d'imprimerie. Une lettre arabe change de forme géométrique en fonction de sa position dans le mot.

Il existe, en général, trois formes pour chaque graphème et ces formes varient selon que le graphème apparaît au début, au milieu ou en fin du mot. Le système d'écriture et celui de l'orthographe de l'AS sont uniformes dans tous les pays arabophones [117]. Trois des 28 lettres de l'alphabet arabe présentent des variantes au niveau de l'écriture. Ces variantes sont au nombre de huit : une variante du *tâ'* (« َ » le [*tamarbu:ta*]), une variante du [*?alif*] dite « brève ou tordue » (« ى » le [*?alifmaqsu:ra*]), cinq variantes du [*hamza*] et le [*?alif mamdu:da*] « ِ » qui est un caractère remplaçant soit la suite « *hamza* voyellé + *hamza* quiescent » soit la suite « *hamza* avec *alif* comme support + *fatha* + [a:] ».

Variantes de consonnes	
<i>tâ' marbûta</i> [<i>tamarbu:ta</i>]	ة
<i>Alif maqsûra</i> [<i>?alifmaqsu:ra</i>]	ى
<i>Alif hamza</i> [<i>?alif hamza</i>] en chef	أ
<i>wâw hamza</i> [<i>wa:whamza</i>] en chef	ؤ
<i>yâ' hamza</i> [<i>ja:? hamza</i>] en chef	ئ
<i>Alif hamza</i> [<i>?alif hamza</i>] souscrit	إ
<i>Alif wasla</i> [<i>?alif wasla</i>] en chef	آ
<i>Alif madda</i> [<i>?alif mamdu:da</i>] en chef	آ

Tableau 2.1: Speech Assessment Methods Phonetic Alphabet (SAMPA) de l'AS (Consonnes, voyelles, diphtongues et quelques symboles)

Graphème de l'AS	Symbole phonémique	Graphème de l'AS	Symbole phonémique
Consonnes			
ء	/ʔ/	ض	/dʔ/
ب	/b/	ط	/tʔ/
ت	/t/	ظ	/Dʔ/
ث	/T/	س	/ʔʕ/
ج	/ʒ/, /dʒ/, /g/	ش	/G/, /ʒ/
ح	/x/	ف	/f/
خ	/X/	ق	/q/
د	/d/	ك	/k/
ذ	/D/, /ð/	ل	/l/
ر	/r/	م	/m/
ز	/z/	ن	/n/
س	/s/	ه	/h/
ش	/S/, /ʃ/	و	/w/
ص	/sʔ/	ي	/j/
Voyelles (courtes et longues)		Diphtongues	
	/a/		/aj/
Voyelle longue Alif	/a:/		/aw/
	/i/	Voyelles casuelles ([tanwi:n])	
Voyelle longue Ya	/i:/	ان	[tanwi:n fátxa] [an]
	/u/	ام	[tanwi:n d'amma] [un]
Voyelle longue Waw	/u:/	ان	[tanwi:n kasra] [in]
Signes de syllabation			
ء	chadda		
و	[su:koun]		

Tableau 2.2 : Phonèmes Consonantiques de l'AS

		Bilab./ Labio- dent.	Ordinaire		Emphatique		Palat.	Vél.	Uvul.	Pharyn g.	Épiglottale	Glott.
			Dent.	Alvéol.	Alvéol.	Dent.						
Nasales		M		n								
Occlusives	Muettes	(p)		t		tʕ		k	q			ʔ
	Sonore	B		d		dʕ	j ~ g					
Fricatives	Muettes	F	θ	s	ʃ	sʕ		x ~ χ		ħ ~ ɦ		h
	Sonore	(v)	ð	z	ʒ ~ dʒ	zʕ ~ ðʕ		ɣ ~ ʁ			ʕ ~ ʔ	
Spirantes		W			l ~ lʕ		j					
Vibrantes					r ~ rʕ							

2.4.2 Problèmes morpho-phonémiques

Comme en Anglais et en Français, la conversion des graphèmes arabes peut dépendre des mots précédents et/ou suivants. En arabe, ce type de dépendance contextuelle est rencontré avec tout mot commençant par le préfixe " ال " (l'équivalent de "la" ou « le » du Français, ou « the » pour l'Anglais) qui est suivie par ce qui est normalement appelé en arabe une « *consonnes solaires* » (Tableau 2.4).

Lorsqu'un mot commence par " ال " suivi par une lettre « consonne solaire » et le mot est précédé par une voyelle, le préfixe 'ال' est prononcé comme /ʔa/, le caractère solaire suivant est géminé (مَشْدَد), et le mot qui contient le préfixe " ال " est confondu (lié) avec son prédécesseur. Par exemple, « إنكسرت الطاولة » [ʔinkasaratiʔat't'a:wilatu] (la table est cassée). Si le mot préfixé par "ال" n'est pas précédé par une voyelle, le " ال " est prononcé comme /ʔa/ et la « lettre solaire » suivante est géminée mais les deux mots ne seront pas liés en prononciation.

Tableau 2.3 : Consonnes de l'AS et leurs correspondants en Français (Si elles existent)

	Occlusives	Emphatiques	Fricatives	Nasales	Liquides	Glides (semi-voyelles)	
Labiales	ب b		ف f	م m		و w	
Interdentales		ظ	ذ	ث			
Dentales	د d	ت t	ض	ط	ن n	ل l	ر r
Sifflantes			ص s	ز z	س s		
Palatales		ج		ش		ي y	
Vélaires		ك k		غ	خ		
Uvulaire		ق q					
Pharyngales				ع	ح		
Glottales		ء		e	ه h		

En effet, cette situation s'appelle « Voyelles d'appui », ou encore « voyelles de liaison ». Les voyelles d'appui sont les trois voyelles brèves (la damma, la fatha et la kasra) utilisées, dans un cadre de conditionnement phonologique (syllabique), pour remplacer un /suku:n/ « ْ » à la fin d'un mot-outil ou d'un verbe suivis d'un mot commençant par un hamza instable, et ce dans le but de résoudre un problème phonologique. En effet, il est impossible, en arabe, d'avoir deux sukoun qui se suivent immédiatement.

Alors, quand on se trouve devant le cas où un mot se terminant par un sukoun (مِنْ [min] « de », مَنْ [man] « qui?, celui qui », هُمْ, « [hum] « eux », قَالَتْ [qa:lat] « elle a dit ») et suivi par un mot commençant par un hamza instable, on remplace le « suku:n » du premier mot par une voyelle d'appui ; cette voyelle est en général une kasra sauf dans les cas suivants :

- une fatha quand il s'agit de مِنْ suivie de l'article ال;
- une damma quand il s'agit des pronoms pluriels masculins هُمْ, كُمْ, ou du suffixe verbal du pluriel masculin ... تُمْ .

Les exemples qui suivent enregistrent bien le remplacement du soukoun dans chacun des cas énumérés (ceci représente le cas général, mais il y a quelques exceptions) :

مِنْ + الْكِتَابِ	→	مِنْ الْكِتَابِ	« du livre »
مَنْ + الْوَلَدِ ؟	→	مَنْ + الْوَلَدِ ؟	« Qui est l'enfant ? »
قَالَتْ + الْبِنْتِ	→	قَالَتْ + الْبِنْتِ	« La fille a dit »
هُمْ + النَّاسِ	→	هُمْ + النَّاسِ	« Ce sont les gens »
ضَرَبْتُمْ + الْكُرَةَ	→	ضَرَبْتُمْ + الْكُرَةَ	« vous avez frappé la balle »

Tableau 2.4 : Classification des consonnes tenant compte des contraintes de la transcription.

Solaires	Lunaires
ت ث د ذ ر ز س ش ص ض ط ظ ل ن	ء ا ب ج ح خ ع غ ف ق ك ه م و ي

2.4.3 Irrégularités orthographiques en AS

L'AS est une langue quasiment phonogrammique dans la mesure où chaque phonème est pratiquement toujours transcrit par un seul graphème : «L'orthographe arabe étant relativement saine, le passage du plan de la graphie à celui de la phonie ne présente pas de grosses difficultés ». Il y a, cependant, certaines variations phonétiques (emphatisation, gémiation, nasalisation, pharyngalisation, assimilation ... et qui seront formalisées au chapitre 3), en plus certains mots présentent une inadéquation entre les graphèmes et les phonèmes. Ce dernier phénomène des irrégularités orthographiques se traduit par plusieurs cas de figure :

- problèmes d'élision (ce qui est écrit mais non prononcé) ;
- ce qui est élidé mais toujours prononcé.

2.4.3.1 Problèmes d'élision (ce qui est écrit mais non prononcé)

L'élision et l'épenthèse sont des problèmes souvent rencontrés en Français [60]. Ils concernent surtout la prononciation du graphème "e" qui est parfois omise et devient un phonème [Φ] ou pareil au son du schwa. En langue arabe, les problèmes de l'élision sont rencontrés avec le graphème "ا" Alif [ʔalif]. Le graphème Alif est parfois omis ou bien réalisé comme étant le son de la voyelle longue [a:]. Si le graphème "ا", est en fin du mot, il n'est pas prononcé, et s'il est au milieu, il sert à vocaliser le graphème précédent en lui ajoutant un son de la voyelle longue [a:]. Par exemple, dans le mot « كتبوا » [katabu:] (ils ont écrit), le graphème « ا » à la fin du mot ne produit aucun son et est omis. Dans le mot "دواب" [dawa:b] (montures) le graphème "ا", est utilisé pour vocaliser le graphème "و" [w] suivi par le son de la voyelle longue [a:].

Le graphème "ا" est également omis quand il apparaît à la fin d'un mot après [tanwi:n fatxa]. Par exemple dans le mot "كبيراً" /kabi:ran/ (énorme), le Alif est omis. Le graphème "ا" peut aussi se produire au milieu comme étant [hamzat wasl] همزة وصل, et dans ce cas, il est omis. Ce [hamza] de liaison (ou hamza instable) est un [hamza] initial qui est écrit (il ne l'est pas systématiquement) avec son support (le alif de l'article) et prononcé quand il commence une phrase (اُكْتُبْ [ʔuktub]), alors qu'il n'est pas prononcé, puisqu'assimilé à la lettre qui le suit, et seulement son support (le alif) est écrit quand il est au milieu d'une phrase, c'est-à-dire quand il est précédé d'un mot ou d'une particule (اِقْرَأْ دَرْسَكَ وَأَكْتُبْ فَرْصَكَ). Pour marquer graphiquement l'assimilation du hamza on utilise, parfois, un signe diacritique spécial écrit au-dessus du alif (آ) et c'est, en fait, ce signe-là que l'on appelle la « wasla ». Comme l'utilisation de ce signe diacritique n'est pas systématique surtout dans les tapuscrits, nous préconisons, dans un souci d'harmonisation, de ne pas saisir la wasla. L'alif, seul, est donc saisi à la place du «alif wasla en chef ». A l'origine, le signe diacritique au-dessus du «alif wasla en chef » est la lettre « ص » qui est l'initiale de l'impératif « صِلْ » (« ici » tu dois faire une liaison). Cette annotation est

empruntée au système d'écriture du Coran et qui s'y trouve pour en faciliter la lecture et la psalmodie التجويد.

En fait, L'écriture du [hamza] constitue le seul problème majeur que l'orthographe arabe connaisse. Le hamza s'écrit soit seul, directement sur la ligne (ء), soit au-dessus ou au-dessous d'un support qui peut être le alif (أ), le wâw (ؤ) ou le yâa sans les points qu'on appelle la nibra (أ – إ) et ce en fonction de la voyelle du hamza et de celle le précédant.

- au début du mot : au début du mot, le hamza a toujours l'alif comme support. Si la voyelle du hamza est une fatha ou une damma, le hamza est écrit au-dessus de l'alif (أ). Si sa voyelle est une kasra, le hamza est écrit au-dessous de l'alif (إ).
- au milieu du mot : les règles d'écriture du hamza sont basées sur un certain « rapport de force » des voyelles qui sont classées, de la plus forte à la plus faible, comme suit : la kasra « ِ », la damma « ُ », la fatha « َ » et en fin le [su:ku:n] « َو » (/i/ > /u/ > /a/ > ∅).

Le hamza est écrit sur le support qui correspond à la voyelle la plus forte entre celle du hamza et celle qui le précède. Toutefois, une seule exception est à noter ici : le hamza n'a aucun support s'il est précédé d'un «[a:]» et porte une fatha (exemple : جاءَ). Le Tableau 2.5 suivant résume ces règles :

Tableau 2.5 : Supports du hamza au milieu du mot

Au milieu du mot				
Voyelle avant le hamza →	kasra « ِ »	damma « ُ »	fatha « َ »	suku:n « َو »
Voyelle du hamza ↓				
kasra « ِ »	ئ	ئ	ئ	ئ
damma « ُ »	ئ	ؤ	ؤ	ؤ
fatha « َ »	ئ	ؤ	أ	أ
suku:n « َو »	ئ	ؤ	أ	ء

- à la fin du mot, le support du hamza dépend uniquement de la voyelle le précédant, comme le montre le Tableau 2.6 ci-dessous :

Tableau 2.6 : Supports du hamza à la fin du mot

A la fin du mot				
Voyelle avant le hamza →	kasra « ِ »	damma « ُ »	fatha « َ »	Suku:n « َو »
	ئ	ؤ	أ	ء

En fait, nous avons tenu à rappeler ici les règles d'écriture du hamza dans le but principal d'insister sur la nécessité de préparer, après la segmentation, la graphie des mots comportant un hamza pour la transcription.

Par exemple, dans le mot « طَأَطَأ » (hocher la tête), les deux Alif ([ʔalif]) contenant le *hamza*, sont considérés comme support du phonème hamza (ou bien on peut les considérer comme étant [hamzat wasl] et sont omis et le mot est prononcé « طَأْطَأْ ».

طَأَطَأ
 ءَ - طَ - ءَ - طَ
 طَ - ءَ - طَ - ءَ
 /t'a ʔa ʔa/

Un autre exemple, dans le mot « بُوْبُو » (Pupille), les deux « waw » ([w]) contenant le hamza, ne sont en fait qu'un support orthographique de celui-ci, et ils sont omis au cours du pré-traitement (Etape d'harmonisation).

بُوْبُو
 ءَ - بَ - ءَ - بَ
 بَ - ءَ - بَ - ءَ
 ↓
 /buʔbuʔun/

En fin un dernier exemple avec le *hamza* sur la *nibra*, avec le mot « مَلَائِكَةٌ » (anges).

مَلَائِكَةٌ
 ءَ - لَ - ءَ - لَ - ءَ - لَ - مَ
 ↓
 /mala:ʔikatun/

Le numéral *cent* (مائة) peut s'écrire avec un *alif* orthographique au singulier, au duel (مائتان) et agglutiné à d'autres numéraux, les unités en l'occurrence, comme dans « ثلاثمائة » (trois cents). Il y a d'autres graphies possibles de ce mot comme : مئة (cette graphie est en fait sa graphie originelle) ou, très rarement مائة.

2.4.3.2 Ce qui est éliminé mais toujours prononcé

- L'*alif* de la particule d'attention «ها» est éliminé quand celle-ci est préfixée aux démonstratifs :

- هذا mais prononcé «هاذا/ha:Da:/» (ce, ceci, celui-ci)
- هذه mais prononcé «هاذه/ha:Dihi/» (cette, celle-ci)
- هؤلاء mais prononcé «هاؤلاء/ha:ʔula:ʔ/» (ceux-ci, celles-ci)

Ou quand celle-ci est préfixée aux pronoms personnels commençant par *hamza* :

- أنا mais prononcé «هاأنا/ha:ʔana:/» (je suis là)
- هأنتم mais prononcé «هاأنتم/ha:ʔantum/» (Vous êtes là)
- هأنتما mais prononcé «هاأنتما/ha:ʔantuma:/» (Vous êtes là « à la forme duale »)

Ou quand celle-ci est également préfixée au pronom «أنا» (le pronom du sujet « Je ») suivi du démonstratif «ذا» هأنذا mais prononcé «هاأنذا/ha:ʔanaDa:/» (Me revoilà !).

- L'alif du démonstratif « ذَا » est élide quand est suffixé à ce dernier un *lâm* comme dans : ذلك « /Da:lika/ » (cela, celui-là), ذلكما « /Da:likuma:/ », « ذلكم » « /Da:likum/ », « ذلكن » « /Da:likunna:/ » (même sens : cela, celui-là ; mais en s'adressant à deux ou plusieurs personnes) ;
- Certains mots sont toujours prononcés avec un *alif* (voyelle longue /a:/) qui n'est pas écrit, à savoir :
 - الله « /ʔalla:h/ » (Allah)
 - إله « /ʔilla:h/ » (un dieu)
 - لكن « /la:kin/ » (mais, devant une phrase)
 - لكنَّ « /la:kinna/ » (mais, devant un nom ou un pronom).
- Les pronoms relatifs de la troisième personne du singulier masculin الذي « /ʔallaDi:/ » et féminin التي « /ʔallati:/ », et de la troisième personne du pluriel masculin الذين, s'écrivent avec une seule lettre *lâm* (celle de l'article) mais se prononcent avec deux *lâm*. Il est à noter que le même relatif au duel et au pluriel féminin, n'est pas dans ce cas, l'adéquation entre graphèmes et phonèmes y est respectée : – اللذان – اللتين – اللاتي – اللتان, ainsi que اللواتي – اللاتي ;
- Le deuxième Waw /w/ « و » (la voyelle longue) dans (داود) /dawu:d/ et (طاووس) /t'awu:s/ est élide par écrit, mais toujours prononcé.

2.4.4 Noms propres et mots étrangers

Les noms propres en langue arabe sont des noms arabes ou étrangers. Les noms d'origine arabe obéissent aux règles de l'orthographe de la langue arabe et sont gérés par les règles lettre-son. Des exemples de noms propres en arabe : « محمد » /muxammad/ (Mohammed), « أحمد » /axmad/ (Ahmed), etc. Un dictionnaire de prononciation peut prendre en charge les mots étrangers, mais ce genre de dictionnaire ne peut être exhaustif. La Morphologie, dans ce genre de situations, pourrait être d'une grande utilité, surtout avec les noms de lieux. D'autres systèmes utilisent les prononciations des mots analogues [102, 105, 99, 101]. En Europe, de grands efforts sont déployés pour résoudre le problème des noms propres européens, en utilisant des dictionnaires unifiés de prononciation [118]. Quand ils apparaissent dans le texte arabe, les noms propres ou des mots étrangers intégrés en leurs langues d'origines, sont souvent en majuscule (ou en italique) et peuvent obéir à des règles lettre-phonème de leurs langues respectives. Ils peuvent être analysés séquentiellement ; puis un petit ensemble de règles (selon une estimation de l'identité de la langue étrangère) pourrait être appliqué à ces mots. Vitale et Belhoula ont développé et utilisé cette approche quand les noms propres et de lieux ou les mots en Anglais sont intégrés dans le texte en AS [119,120].

Le problème en arabe standard de néologisme ou l'émergence de nouveaux mots est causé soit par l'évolution linguistique de la langue arabe ou soit par l'influence d'une langue étrangère dominante, comme l'Anglais ou le Français, ce qui provoque l'émergence de nouveaux mots, ou en raison du progrès technologique qui stimule l'introduction de nouveaux mots ou termes techniques. L'invention de la voiture, de l'avion, de la fusée ou de l'ordinateur et l'introduction de ces mots dans les dictionnaires Anglais et Français, a incité les linguistes arabes à trouver les mots arabes équivalents. L'approche favorable adoptée par ces linguistes était d'introduire de nouveaux mots arabes équivalents aux mots Anglais et Français, sur la base de la fonctionnalité de l'appareil inventé.

Le mot arabe utilisé pour «ordinateur», par exemple, est «الحاسوب» qui est basé sur sa fonctionnalité de calcul. Une autre approche désagréable pour les linguistes arabes est d'inventer une adaptation du mot étranger nouvellement introduit à une prononciation qui correspond à la structure des mots arabes. Le mot arabe utilisé pour «télévision», par exemple, est «تلفاز» qui adapte simplement la prononciation Anglaise du mot à une forme maniable en arabe. En outre, l'utilisation des noms propres Anglais ou Français est également une pratique courante en arabe et la prononciation de ces noms, dans la plupart des cas, est maintenu dans sa forme originale en Anglais ou en Français. Pour ces cas, il est plus judicieux de faire appel aux systèmes TTS Anglais ou Français en collaboration avec le système TTS arabe pour fournir une solution partielle au problème. Cette approche ne sera pas adoptée dans notre présente étude, mais fera l'objet d'un projet de recherche annexe dont les résultats seront exposés dans une future publication.

2.4.5. Syllabation

La Syllabation en conversion TTS est importante pour deux raisons. Tout d'abord, elle permet la mise en œuvre de certaines règles lettre-phonème comme la génération de diphtongue dans la Conversion Graphème-Phonème ou le traitement de la pharyngalisation dans la conversion phonème-phone. Deuxièmement, la syllabation est essentielle dans l'amélioration de la qualité de la parole produite par des synthétiseurs puisque la détection de la syllabe aidera à modéliser les durées des phones et est utilisée pour se renseigner sur certains traits acoustiques comme l'intensité et la durée pour améliorer l'intonation de la parole à synthétiser [121].

En AS, il existe six types de syllabes : [CV], [CV:], [CVC], [CV:C] et, les deux rares formes, [CVCC] et [CV:CC] où V représente une voyelle courte et V: signifie une voyelle longue. Les trois premiers types sont des types de syllabes les plus courantes en langue arabe. Al-Ani [122], Mitchell et Anis [123] décrivent amplement les syllabes de la langue arabe. La séquence phonémique peut être divisée en syllabes en sachant que le noyau de chaque mot en langue arabe standard est une voyelle. Cette propriété est très intéressante et utile pour extraire les syllabes d'une représentation phonémique de mots arabes. Chaque mot phonémique est analysé de manière descendante (un parcours de droite à gauche à partir de la fin du mot phonémique) à la recherche de voyelles. Chaque fois qu'une voyelle est détectée, une autre analyse plus localisée est effectuée à la recherche d'une correspondance syllabique avec l'un des six types de syllabes indiqués ci-dessus. L'algorithme de syllabation adoptée dans cette étude sera présenté au chapitre 3.

2.4.6 Normalisation du texte et pré-traitement

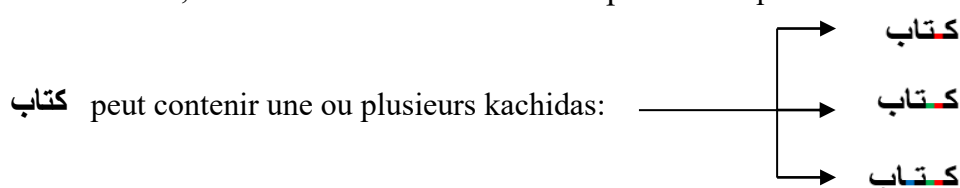
La normalisation ou la mise en forme et le pré-traitement du texte est une interface essentielle pour tout système produisant des sons de la parole à partir du texte écrit. Le pré-traitement de texte est nécessaire pour préparer l'entrée des différents traitements et analyses pris en charge par les modules restants du système TTS. Les tâches faisant partie du processus de pré-traitement sont :

- La Segmentation du texte ou le partage du texte en unités lexicales distinctes bien formatés, telles que des mots. En effet, le terme segmentation ne doit pas être compris ici dans le sens de la segmentation lexicale, mais dans le sens de l'opération qui consiste à délimiter des unités minimales dans un texte. Il s'agit tout simplement de quantification, c'est-à-dire du passage du texte aux listes. Cette segmentation opère donc au niveau des formes graphiques, comprises entre les délimiteurs, « les mots ».
- Les Conversions des acronymes, abréviations et des caractères ou des symboles non-alphanumériques en descripteurs appropriés de mots ou de phrases. Le caractère non-alphanumérique %, par exemple, est convertit en séquence de mots arabes " في المائة " (pour cent), ou l'abréviation « ذ م م » en séquence de mots arabes « ذو مسؤولية محدودة » (à responsabilité limitée) (Voir *Tableau 3.12* du chapitre 3).

- *Les expressions numériques telles que* : nombres, poids et mesures, dates et heures, fractions... sont omniprésentes dans les textes en langage naturel et ont généralement, une grammaire irrégulière idiosyncrasique. Cependant, comme elles sont construites avec quelques articles, déterminants et constructions lexicales bien définis, elles peuvent être assez facilement identifiées. Dans notre étude, nous essayons de concevoir et de mettre en œuvre un système de détection et d'identification de telles expressions dans les textes arabes afin de les traduire en transcription orthographique régulière (en séquences de mots appropriés) qui précède la transcription phonétique globale du texte. La tâche centrale de cette opération est le repérage et l'identification de toutes les expressions numériques qui peuvent apparaître dans le texte et convertir ces chiffres en mots en langue arabe standard, ce que nous avons implémenté en *annexe 04*. Nous comptons développer – dans un futur projet – un système automatique de repérage des expressions numériques fondé sur une analyse au moyen d'algorithmes génétiques, puis de convertir ces objets en mots, de sorte que :
 - la conversion soit grammaticalement correcte pour les nombres entiers de 0 à 10²¹ et plus ;
 - des expressions soient sensibles au genre (masculin et féminin (مذكر ومؤنث)) ;
 - la conversion répondant à la spécificité de la grammaire arabe concernant les formes : nominative, accusative et génitive (رفع ، جر ، ونصب).

Cette approche est originale, et offre une base nécessaire à de nombreuses applications. Comme exemples : La conversion des dates « 21-05-2001 » (21/05/2001) en une séquence de mots appropriés, la conversion de nombres fractionnaires (1/2, 1/4) et tous les nombres (256) ou les nombres entiers avec parties fractionnaires (256.8) en séquences de mots appropriés.

- Traitement de la kachida ou [tatwi:l] : la kachida ou encore « حرف تطويل » (*harf tatwi:l* : un symbole d'allongement) est introduite dans les mots en arabe pour augmenter la distance entre certains caractères à l'aide d'un trait «-» (obtenu en combinant, en clavier arabe, « shift+ت »). Elle est souvent rajoutée pour des raisons purement esthétiques. De nos jours, avec les textes électroniques, les utilitaires de traitements de textes ont recours à la Kachida pour l'ajustement des paragraphes arabes par le biais d'une augmentation des espaces entre les caractères en vue d'une meilleure visibilité. Exemple : prenons le mot « تتشبتون » /tataSabbaTu:na/ (vous tenez à qqchse ou à qq'un), la séquence est plus lisible quand elle est notée ainsi « تتشبتون » /ta_ta_Sa_bba_Tu:_na/. L'emplacement de la Kachida dans le mot est très aléatoire. Or, ce caractère ne figure pas dans l'alphabet arabe et par conséquent n'est pas concerné par la transcription orthographique-phonique. Les méthodes traditionnelles de traitement automatique du texte commencent par parcourir le mot pour éliminer la Kachida avant d'entamer toute opération de pré-traitement (segmentation, harmonisation, ...) [124]. En effet, si les kachidas ne sont pas traitées en amont, elles seront conservées lors de la segmentation et des étapes suivantes. Ceci va engendrer l'éparpillement des formes des graphies sur plusieurs vocables alors qu'elles ne sont, en réalité, que les occurrences d'un même et unique vocable. On peut, en effet, avoir la même forme écrite tantôt sans aucune kachida, tantôt avec une seule, tantôt avec deux kachidas ou plus. Exemple :



- Il est important de noter ici qu'il ne faut pas saisir le [tanwi:n] au-dessus du *alif* orthographique comme le font, à tort, certains imprimeurs dans certains pays arabes. Il faut, au contraire, suivre la tradition typographique arabe en saisissant le [tanwi:n] avant le *alif* orthographique, c'est-à-dire au-dessus de la consonne qui précède ce dernier. Pour cela, au cours du pré-traitement du texte, une normalisation de cette situation s'impose si elle est rencontrée. Exemple :

كبيراً ➡ كبيرًا

Le pré-traitement commence par segmenter le texte d'entrée en mots et en phrases. McAllister dans [125] a proposé de diviser le texte en îlots orthographiques (chaînes de caractères ASCII délimités par des caractères d'espace blanc, un espace, un caractère de tabulation ou de retour nouvelle ligne). Pour éviter les ambiguïtés liées à l'intégration des signes de ponctuation dans les îlots orthographiques, les signes de ponctuation de la langue arabe sont isolés et plus tard seront réaffectés aux îlots orthographiques auxquels ils appartiennent. Il en résulte des unités de segmentation de base. Des simples règles habituelles sont ensuite utilisées pour convertir les unités de segmentation de base en unités de segmentation finale en examinant chaque unité de segmentation de base entrante, dans un ordre de parcours gauche-droite. Quand il s'agit de détecter la fin des phrases, plusieurs ambiguïtés de ponctuation surgissent liées aux marques de fin des phrases (le point (.), le point d'interrogation (?), les deux points (:), la virgule (,) et le point d'exclamation (!)). Le point d'interrogation est presque sans ambiguïté, mais les autres signes sont relativement ambiguës quant à leur utilisation dans le texte. En indiquant l'heure, le signe de deux points (:) est utilisé pour séparer les heures des minutes. Le point d'exclamation est utilisé comme le symbole factoriel en mathématiques. En arabe standard, comme en Anglais ou en Français, par exemple, l'espace insécable et la virgule sont utilisées dans les nombres formés d'entiers et de parties fractionnaires : l'espace insécable est utilisé en Français pour segmenter un grand nombre, alors qu'en Anglais et en arabe standard, la virgule est utilisée pour le même but. En Français, la virgule est utilisée comme un point décimal. Ces types d'ambiguïtés sont sensibles au contexte et sont presque universelle pour toutes les langues utilisant un système de ponctuation similaire à l'Anglais. Liberman et Church ont largement discuté ces ambiguïtés et ont proposé une approche probabiliste pour l'ensemble du problème de segmentation du texte [126].

L'approche pour la segmentation du texte pour la langue arabe en phrases bien structurées puis en de simples mots suit les propositions décrites ci-dessus. Les problèmes de base tels que la tokenisation en mots et la détection de fins de phrases simples, le formatage de mots, la consultation du lexique et la conversion des nombres ont été abordées dans [127]. Certains des ambiguïtés concernant la fin des phrases sont sensibles au contexte, surtout dans des domaines spécifiques. Par exemple, l'utilisation du point comme séparateur décimal dans les nombres à parties fractionnaires. En effet, en Français, dans l'expression d'une quantité, les nombres s'écrivent par tranches de trois chiffres séparées par un espace insécable que ce soit pour la partie entière ou décimale : « 12 457 892,561 41 ». Si le nombre n'a que quatre chiffres, les deux écritures sont acceptées, avec ou sans espace « 1 987 » ou « 1987 ». Dans l'expression d'un rang (numéros d'adresses, d'articles, d'années, etc.) on ne met pas d'espace entre les tranches de trois chiffres : « 1245 boulevard Saint-Laurent », « année 2007 ». En Anglais, les mêmes règles sont observées, mais en utilisant la virgule au lieu de l'espace dans la séparation des tranches de trois chiffres dans les grands nombres, et le point au lieu de la virgule pour séparer la partie fractionnaire de la partie entière, Il en est de même pour l'AS.

Un lexique spécial de séquences de mots est utilisé pour maintenir : les abréviations et les symboles usuels en langue arabe, les acronymes et les mots à orthographes irrégulières. Chaque entrée dans le lexique est divisée en deux champs, l'orthographe de l'article et sa séquence de mots correspondante en langue arabe (pour les acronymes, cela peut être la prononciation de l'acronyme ou sa séquence de texte correspondante).

Les algorithmes pour la tokénisation du texte d'entrée, l'expansion des abréviations et des symboles, la détection et le traitement des mots à orthographe irrégulière à partir d'un lexique spécifique, et la prononciation des nombres sont très similaires à ceux utilisés pour l'Anglais et peut-être pour d'autres langues comme le Français. La tokénisation ne sera pas abordée dans cette étude, tout en sachant que la prise en charge de ce point améliore sensiblement la performance du système de conversion lettre-son. Dans le cadre des opérations de pré-traitement, nous pouvons citer ces quelques opérations supplémentaires d'harmonisations qui améliorent l'orthographe du texte, et le prépare à la procédure de Conversion Graphème-Phonème :

2.4.6.1 Chiffres et expressions numériques en forme indienne

Contrairement à ce qui est répandu, même dans certains pays arabes, les chiffres arabes sont ceux utilisés en occident et dans tout le Maghreb « 1, 2, 3, ... ». Les chiffres qu'on utilise dans certains pays arabes du Moyen-Orient « ١, ٢, ٣, ... », sont en fait des chiffres indiens :

Chiffre arabe/Indien	1 / ١	2 / ٢	3 / ٣	4 / ٤	5 / ٥	6 / ٦	7 / ٧	8 / ٨	9 / ٩	0 / ٠
En lettres	Un	Deux	Trois	Quatre	Cinq	Six	Sept	Huit	Neuf	Zéro

2.4.6.2 Ligatures esthétiques

Utilisées dans un but exclusivement esthétique, ces ligatures sont des formes calligraphiques représentant des phrases entières dont les éléments sont entrelacés d'une façon très harmonieuse. Elles ne sont donc pas indispensables sous cette forme et doivent, de ce fait, être remplacées par leurs composantes, non-ligaturées sans changer la validité grammaticale ou le sens du texte à traiter automatiquement. Exemple :

Caractère	Doit être remplacé par :	Traduction
عيد سعيد	عيد سعيد	« Joyeuse fête »
كل عام وأنتم بخير	كل عام وأنتم بخير	« Bonne année »
بسم الله الرحمن الرحيم	بسم الله الرحمن الرحيم	« Au nom de Dieu, clément et miséricordieux »
الحمد لله	الحمد لله	« louange à Dieu »,
الله أكبر	الله أكبر	« Dieu est le plus grand »
صلى الله عليه وسلم	صلى الله عليه وسلم	« Paix et salut d'Allah sur lui »

2.4.6.3 Ponctuation esthétique

Dans certains textes arabes, comme le texte coranique ou autres, on peut trouver des signes de ponctuation sous des formes esthétiques différentes de la forme conventionnelle et ont donc des codes machine différents du code usuel. Il est donc nécessaire de remplacer ces formes esthétiques par les formes usuelles. Les parenthèses, par exemple, du type « ﴿ ﴾ » ou « { } » doivent être remplacées par « () » ou par des guillemets « " » quand ils renferment une citation coranique.

2.4.6.4 Chiffres esthétiques

Comme les signes de ponctuation, on peut trouver dans certains textes arabes, comme le texte coranique, des chiffres sous des formes esthétiques non conventionnelles et ayant des codes machine non-usuels. Il est donc nécessaire de remplacer ces chiffres esthétiques par les formes usuelles. Les chiffres, par exemples, du types « ①, ②, ③ ... » ou « ①, ②, ③ ... » doivent être remplacés par « 1, 2, 3... ». Il est nécessaire de noter ici que la saisie directe des chiffres ou le remplacement des chiffres esthétiques doit se faire d'une façon homogène en utilisant soit les

chiffres arabes (ce que nous suggérons fortement) soit les chiffres indiens et non pas les deux. On pourrait, le cas échéant, faire le choix de convertir, au moment de l'harmonisation, ces chiffres en lettres comme nous le préconisons plus loin dans ce *chapitre* [128].

2.4.6.5 Signes de ponctuation

Dans le but de faciliter l'opération de segmentation, tous les signes de ponctuation doivent être précédés et suivis d'un espace. Ceci évitera d'avoir des cas de figures où l'on a des formes graphiques représentant des lexèmes liés à des virgules, points ou autres. Il est donc préférable d'insérer avant chaque signe de ponctuation un espace. Chaque signe de ponctuation sera ainsi encadré par deux espaces.

2.4.6.6 Mots-outils

Les mots outils sont des entités qui servent à situer des faits ou des objets par rapport au temps ou au lieu. Ils jouent également un rôle clé dans la cohérence et l'enchaînement d'un texte. Par exemple, nous avons des particules qui désignent un temps : بعد (/ba?'da/ - après), قبل (/qabla/ - avant), منذ (/munDu/ - depuis) ou un lieu tel que حيث (/xajTu / - où), etc. Selon leur sémantique et leur fonction dans le texte, ils incluent différentes catégories, nous citons :

- les prépositions : par exemple «في» (/fi:/ - dans) ou «على» (/?'ala:/ - sur);
- les conjonctions de coordination : par exemple, «ثم» (/Tumma/ - ensuite, puis);
- les adverbes : par exemple, «أبدًا» (/?'abadan/-jamais) ou «بشكلٍ عادي» (/biSaklin ?'a:dijjin/ - de façon normale);
- les qualificateurs : par exemple, «كُلُّ» (/kulla/ -tout) ou «بعض» (/ba?'d'a/- un peu); etc.

Ces mots-outils forment donc une liste fermée et ils représentent tout de même un grand pourcentage de l'ensemble des occurrences des textes arabes, il est important, et en même temps aisé, de dresser le tableau de toutes les formes pour chaque mot-outil et de choisir, parmi elles, en toute connaissance de cause la graphie d'harmonisation la plus adéquate. Le *Tableau 2.7* ci-dessous présente une tentative d'inventaire et d'harmonisation de ces éléments.

Tableau 2.7 : Liste des mots-outils et leurs harmonisations

Différentes formes	Graphie d'harmonisation
أ / أُ	أ
إِذَا / إِذَا / إِذَا / إِذَا	إِذَا
إِذَا / إِذَا / إِذَا / إِذَا	إِذَا
إِذَنْ / إِذَنْ / إِذَنْ / إِذَنْ / إِذَنْ / إِذَنْ / إِذَنْ / إِذَنْ	إِذَنْ
إِلَّا / إِلَّا / إِلَّا / إِلَّا / إِلَّا / إِلَّا / إِلَّا / إِلَّا	إِلَّا
الآن / الآن	الآن
أما / أما / أما / أما	أما
أما / أما / أما / أما / أما / أما / أما / أما	أما
أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ	أَنْ
أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ / أَنْ	أَنْ
إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ	إِنْ
إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ / إِنْ	إِنْ
أنا / أنا / أنا / أنا	أنا
أنت / أنت / أنت / أنت / أنت / أنت / أنت / أنت	أنت
أنت / أنت / أنت / أنت / أنت / أنت / أنت / أنت	أنت
أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ	أَيُّ
أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ / أَيُّ	أَيُّ

Différentes formes	Graphie d'harmonisation
أَيْضًا / أَيْضًا / أَيْضًا / أَيْضًا / أَيْضًا / أَيْضًا / أَيْضًا / أَيْضًا	أَيْضًا
...ب / ...ب	...ب
بُئْسَ / بُئْسَ / بُئْسَ / بُئْسَ / بُئْسَ / بُئْسَ / بُئْسَ / بُئْسَ	بُئْسَ
بَعْدَ / بَعْدَ / بَعْدَ / بَعْدَ / بَعْدَ / بَعْدَ / بَعْدَ / بَعْدَ (من) بَعْدَ / (من) بَعْدَ / (من) بَعْدَ / (من) بَعْدَ	بَعْدَ
بَعْدُ / بَعْدُ / بَعْدُ / بَعْدُ / بَعْدُ / بَعْدُ / بَعْدُ / بَعْدُ	بَعْدُ / مِنْ بَعْدُ
مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ	
مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ	
مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ	
مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ / مِنْ بَعْدُ	
بَلْ / بَلْ / بَلْ / بَلْ / بَلْ / بَلْ / بَلْ / بَلْ ...	بَلْ
بَيْنَ / بَيْنَ / بَيْنَ / بَيْنَ / بَيْنَ / بَيْنَ / بَيْنَ / بَيْنَ (من) بَيْنَ / (من) بَيْنَ / (من) بَيْنَ / (من) بَيْنَ	بَيْنَ
تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ	تَمَّ
تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ / تَمَّ	تَمَّ
حَتَّى / حَتَّى / حَتَّى / حَتَّى / حَتَّى / حَتَّى / حَتَّى / حَتَّى	حَتَّى
حَلَوْ / حَلَوْ / حَلَوْ / حَلَوْ / حَلَوْ / حَلَوْ / حَلَوْ / حَلَوْ حَيْثُ / حَيْثُ / حَيْثُ / حَيْثُ / حَيْثُ / حَيْثُ / حَيْثُ / حَيْثُ	حَلَوْ حَيْثُ
حِينَ / حِينَ / حِينَ / حِينَ / حِينَ / حِينَ / حِينَ / حِينَ	حِينَ
عَنْ / عَنْ / عَنْ / عَنْ / عَنْ / عَنْ / عَنْ / عَنْ ...	عَنْ
عِنْدَ / عِنْدَ / عِنْدَ / عِنْدَ / عِنْدَ / عِنْدَ / عِنْدَ / عِنْدَ غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ / غَيْرَ	عِنْدَ غَيْرَ
...ذ / ...ذ	...ذ
فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ	فَوْقَ
فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ / فَوْقَ	فَوْقَ

Différentes formes	Graphie d'harmonisation
فِيمَ / فِيمِ / فِيمَا / فِيمَنْ	فِيمَ
قَدْ / قَدَّ / قَدِّ / قَدِّ	قَدْ
قَطَّ / قَطُّ / قَطُّ / قَطُّ	قَطُّ
كَ / كَ	كَ
كَ / كَ	كَ
كَ / كَ	كَ
كُلَّ / كُلاً / كُلاً / كُلاً / كُلاً / كُلاً / كُلاً / كُلاً	كُلُّ
كَمْ / كَمَّ / كَمَّ / كَمَّ	كَمْ
كَمْ / كُمَّ / كُمَّ / كُمَّ	كُمَّ
كَمَا / كَمَّا / كَمَّا / كَمَّا	كَمَا
كَمَا / كُمَّا / كُمَّا / كُمَّا	كُمَّا
كَيْفَ / كَيْفًا / كَيْفًا / كَيْفًا / كَيْفًا / كَيْفًا / كَيْفًا / كَيْفًا	كَيْفًا
ل... / ل... / ل... (أه - لك - لنا - ...)	ل...
ل... / ل...	ل...
لَقَدْ / لَقَدَّ / لَقَدِّ / لَقَدِّ / لَقَدِّ / لَقَدِّ / لَقَدِّ / لَقَدِّ	لَقَدْ
لَكِنْ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ	لَكِنْ
لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ / لَكِيْنُ	لَكِيْنُ
لَمْ / لَمَّ / لَمَّ / لَمَّ	لَمْ
لَمْ / لَمَّ / لَمَّ / لَمَّ	لَمْ
لَمَّا / لَمَّمَا / لَمَّمَا / لَمَّمَا	لَمَّا
لَوْ / لَوَّ / لَوَّ / لَوَّ / لَوَّ / لَوَّ / لَوَّ / لَوَّ	لَوْ
مَعَ / مَعَّ / مَعِّ / مَعِّ	مَعَ
مِمَّا / مِمَّا / مِمَّا / مِمَّا / مِمَّا / مِمَّا / مِمَّا / مِمَّا	مِمَّا
مِمَّنْ / مِمَّنْ / مِمَّنْ / مِمَّنْ / مِمَّنْ / مِمَّنْ / مِمَّنْ / مِمَّنْ	مِمَّنْ
مِنْ / مَنَّ / مَنَّ / مَنَّ / مَنَّ / مَنَّ / مَنَّ / مَنَّ	مِنْ

Suite et fin (Tableau 2.7) :

Différentes formes	Graphie d'harmonisation
من / مِنْ / مِن / مِنْ / مِنْ ...	مِنْ
مند / مُنْدُ / مُنْدُ / مُنْدُ / مُنْدُ / مُنْدُ / مُنْدُ / مُنْدُ	مَنْدُ
نحو / نَحْوُ / نَحْوُ / نَحْوُ / نَحْوُ / نَحْوُ / نَحْوُ / نَحْوُ	نَحْوُ
نعم / نَعَمُ / نَعَمُ / نَعَمُ / نَعَمُ / نَعَمُ / نَعَمُ / نَعَمُ	نَعَمُ
نعم / نِعَمُ / نِعَمُ / نِعَمُ / نِعَمُ / نِعَمُ / نِعَمُ / نِعَمُ	نِعَمُ
ه / هُ / هِ	ه
هم / هُمُ / هُمُ / هُمُ / هُمُ / هُمُ / هُمُ / هُمُ	هَمُ
هما / هُمَا / هُمَا / هُمَا / هُمَا / هُمَا / هُمَا / هُمَا	هَمَا
هنَّ / هُنَّ / هُنَّ / هُنَّ / هُنَّ / هُنَّ / هُنَّ / هُنَّ	هَنَّ
ههنا / هَهُنَا / هَهُنَا / هَهُنَا / هَهُنَا / هَهُنَا / هَهُنَا / هَهُنَا هاهنا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا هاهنا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا / هَاهُنَا	هَهُنَا
ههيات / هَهُيَاتُ / هَهُيَاتُ / هَهُيَاتُ / هَهُيَاتُ / هَهُيَاتُ / هَهُيَاتُ / هَهُيَاتُ ههيات / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ ههيات / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ / هِهْيَاتُ	هِهْيَاتُ
و / وَ	و

2.5 Conclusion

Nous avons exposé dans ce chapitre les questions liées à la conversion de texte en AS en entités phonémiques et phonétiques pour être ensuite utilisées dans la TOP. Nous avons étudié également les différentes approches de la TOP. Cette analyse est indispensable pour le développement de notre formalisme de la transcription qui sera présenté dans le chapitre suivant.

Chapitre 3 :
TOP-AS : Formalisme, algorithmes,
ressources numériques
et Implémentation

3.1 Introduction

Ce chapitre décrit amplement l'approche à base de règles que nous avons retenue pour notre système TOP (baptisé TOP-AS): Ses fondements, ses méthodes, ses ressources et algorithmes, ainsi que son implémentation.

Ces règles ne sont pas, en général, une correspondance une-à-une (relation biunivoque), mais sont des règles de réécriture contextuelles d'un formalisme similaire à celui de la Phonologie Générative comme dans l'ouvrage de Chomsky et Halle [32]. En effet, au début des travaux sur la phonologie générative, comme a expliqué Chomsky et Halle, les formes sous-jacentes de morphèmes sont constituées de chaînes de « traits phonétiques » abstraites, mais pas arbitraires. Ces caractéristiques sont universelles car elles "représentent les capacités phonétiques de l'Homme» [32]. D'autres développements de la théorie, comme les travaux sur les bases articulatoires de ces traits, ainsi que le rôle phonétique en phonologie, ont reçu une attention considérable, ce qui a abouti ultérieurement (vers 1983) au fait que ces traits phonologiques représentent des instructions neuronales aux articulateurs. Les règles graphème-phonème, adoptées dans notre étude, pour la transcription de texte arabe fonctionnent sur un seul niveau, le niveau des graphèmes, et produisent un niveau de sortie contenant les phonèmes.

Ces règles sont de la forme (C'est Une écriture possible parmi plusieurs d'autres, comme nous allons voir dans la section suivante.) :

$$A \longrightarrow [B]/ X_Y;$$

Où A et B peuvent être un caractère orthographique unique, des chaînes de caractères, ou nul. La règle ci-dessus signifie que « **le graphème A devient le phonème [B] si A est entre le contexte graphémique gauche X et le contexte graphémique droit Y** », c'est-à-dire, une lecture de droite à gauche, de la façon suivante :

$$[\text{Phonème}] = \{\text{CG}(\text{contexte gauche})\} + \{\text{C}(\text{caractère})\} + \{\text{CD}(\text{contexte droit})\}$$

Il est important de noter que l'exécution consécutive de règles peut conduire à des conflits si plus d'une règle est applicable à un certain point. Les conflits se produisent si l'application d'une règle produit une chaîne de sortie à laquelle une autre règle pourrait alors s'appliquer ou si l'application d'une règle consomme les lettres dans le niveau graphème qui autrement auraient déclenché d'autres règles. Il est important de noter que les deux niveaux d'entrée et de sortie des règles graphème-phonème sont différents. Le niveau d'entrée contient les graphèmes et celui de sortie, les phonèmes. Etant donné que les niveaux d'entrée et de sortie sont différents, le premier type de problèmes ne se pose pas, mais le second pourrait surgir. Un ordonnancement minutieux des règles pourrait le réduire considérablement.

En appliquant les règles que nous allons exposer plus bas, nous avons jugé utile de transformer les groupes de consonnes avant les voyelles afin que leurs sons fournissent des informations de contexte supplémentaires pour transformer les voyelles.

3.2 Formalisme du système TOP-AS

Les règles de réécriture peuvent être de plusieurs niveaux (appelés règles de réécriture à plusieurs niveaux, MLRRs = Multi Level Rewrite Rules). Ces règles MLRRs utilisent les contextes gauche et droit du graphème en cours d'analyse dans le niveau d'entrée (dans ce cas, le niveau graphème) ainsi que les valeurs associées au graphème dans d'autres niveaux tels que le niveau des phonèmes. Ceci permet de limiter l'application de certaines règles à une séquence particulière de catégories syntaxiques ou restreindre l'application de certaines règles en ajoutant des informations contextuelles issues d'autres niveaux [4, 129]. Les MLRRs sont de la forme :

$$A \longrightarrow [B]/ X_Y/ \text{niveau } i : X_i_Y_i/ \text{niveau } j : X_j_Y_j \dots / \text{niveau } k : X_k_Y_k.$$

La règle MLRR ci-dessus énonce simplement que si A se trouve dans le niveau d'entrée (niveau graphème) entouré par X et Y, et par $X_i_Y_i$, $X_j_Y_j$, . . . , $X_k_Y_k$ aux autres niveaux ($i, j, \dots; K$), le B devrait être produite sur le niveau de sortie. Les autres niveaux peuvent être les niveaux de phonèmes ou de syllabation. Les MLRRs fournissent des solutions au problème de la consommation de graphèmes causée par l'application de certaines règles, que nous avons décrits ci-dessus. Mais un ordonnancement prudent des règles comme indiqué ci-dessus pourrait résoudre ce problème. Extraire des caractéristiques de niveau supérieur pour écrire les MLRRs n'est pas une tâche facile. En outre, l'application des MLRRs est compliquée par le fait que les détails linguistiques supplémentaires sont nécessaires, tels que l'ordonnancement des règles, la direction de parcours de la règle (de gauche à droite ou de droite à gauche) et la hiérarchie de parcours (des petites unités à des unités plus grandes ou vice versa) [4]. Néanmoins, les MLRR ont été appliquées avec succès pour la transcription graphème-phonème des langues complexes comme le Français. Il est assez intéressant de savoir que les MLRRs ont également été utilisées pour la transcription de l'arabe dans le cadre du projet MBROLA, Euler [130]. Nous aurions souhaité dresser une comparaison rigoureuse entre les règles graphème-phonème utilisées pour l'arabe dans le projet Euler et celles que nous avons adopté dans cette étude, mais rien n'a été rendu public concernant le jeu de règles utilisé dans le projet Euler pour la langue arabe.

3.2.1 Règles phonémiques de l'AS

Les règles Graphème-Phonème et Phonème-Phone que nous avons adopté vont nous permettre une phonétisation automatique du texte écrit en arabe. Des connaissances linguistiques et phonétiques sont largement exploitées pour l'élaboration de ces règles. Les règles Graphème-Phonème opèrent sur l'entrée orthographique pour créer les sons de base appropriés de la langue arabe standard. Ces règles associent à chaque séquence de caractères orthographiques une chaîne de phonèmes. Une complication concernant la terminologie de ce domaine est le fait de considérer le graphème comme étant une unité minimale de l'orthographe plutôt qu'un seul caractère, de sorte que dans le mot Anglais « *The* » le « *th* » est décrit comme un seul graphème, et la même remarque pour le Français pour les sons « *ch* » ou « *kh* » etc. Comme l'entrée est toujours une séquence de caractères, c'est la conversion caractère-phonème qui est réellement effectuée pour ces langues. Cet embarras est absent en AS. Cependant, dans la synthèse de la parole et la reconnaissance de la parole, le post-traitement phonétique est souvent nécessaire pour convertir les phonèmes en une séquence de phones ou d'allophones. L'entrée à ce stade est la chaîne phonémique générée par l'application des règles phonémiques.

Le système orthographique de l'AS a été largement abordé dans la littérature à des fins de formalisation. Les contributions les plus riches de la littérature sur l'orthographe arabe peuvent être trouvées dans Hassanein et Shahata [94], Qazzi [95] et Humoud [131]. L'orthographe arabe est assez régulière, sauf dans certaines situations où certains mots violent les règles strictes de prononciation comme nous l'avons détaillé dans le chapitre précédent. Dans ce qui suit, nous allons formaliser les énoncés de ces irrégularités génériques, avec un bref rappel de chacune d'elles. Les exceptions à l'orthographe régulière de la langue arabe se manifestent par son échec en trois catégories dominantes :

- les violations à la génération de la voyelle longue /a:/ ou règle /mad bil?alif/ (voir règle (9) ou «←») (fatha)→[a:]/X_Y, X = {tout graphème qui représente une consonne} et Y = {"|"}). La règle /mad bil?alif/ est violée pour de nombreux mots, qui sont principalement les pronoms démonstratifs. Les exemples sont des mots comme "هذا" /ha:Da:/ (*ce*), " هؤلاء " / ha:?ula:/ (*ces*, le démonstratif au pluriels pour les personnes), etc.
- les abréviations, les symboles et les acronymes ;
- Certains noms exceptionnelles tels que les noms d'Allah " الله /?alla:h/, « الرحمن » /alraxma:n/ et les noms composés dérivés des noms d'Allah comme « عبد الله » /?'abdalla:h/, « عبد الرحمن » /?'abdarraxma:n/. Les mots et les noms exceptionnels sont extraits des plus célèbres et des plus complets des lexiques de mots et de noms arabes, Al Qamous Almuhiyet « القاموس المحيط » [132].

Les mots exceptionnels - qui ne peuvent être énoncés sous forme de règles- sont placés dans un lexique de prononciation des exceptions. La forme phonémique de chaque mot exceptionnel est inscrite en face de sa forme graphémique. Lors de l'analyse de texte, le dictionnaire d'exceptions est consulté en premier avant le traitement des règles.

Pour l'AS, l'ensemble de règles phonémiques comprend 11 catégories de règles graphème-phonème et chaque catégorie a ses propres sous-règles. Les 11 catégories sont comme suit :

(1)- **La règle du [suku:n]**. En arabe standard, « ْ » (suku:n) est un caractère de silence, c'est à dire qu'il n'a pas de prononciation et est normalement omis. La règle suku:n est de la forme :

$$\begin{aligned} \text{« ْ »} &\longrightarrow [\Phi]/ X_Y; \\ X, Y &= \{\text{tout graphème représentant une consonne}\}, \\ \Phi &= \text{phonème nul ou graphème nul} \end{aligned}$$

(2) **La règle d'élision**. Il existe quatre sous-règles dans cette catégorie. Ces règles s'appliquent à « ا » (Alif) quand il se produit en fin ou au milieu d'un mot. Nous avons déjà décrit ces règles dans la *section 2.4.3* du chapitre précédent.

a- Lorsqu'un Alif final « ا » dans un mot (c'est à dire se produisant en fin de mot) est précédé par « و » « w », la combinaison signifie que le mot est au pluriel. Dans ce cas, le dernier « ا » (Alif) est redondant et n'est pas prononcé. La règle prend la forme :

$$\begin{aligned} \text{« ا »} &\longrightarrow [\Phi]/ \text{« و »} \text{« w »} _Y; \\ Y &= \{\Phi, \text{graphème nul}\}. \end{aligned}$$

Par exemple, dans le mot pluriel « كتبوا » (ils ont écrit) le (Alif) final « ا » est omis et le mot est prononcé comme /katabu:/.

b- Lorsque le (Alif) se produit en fin de mot mais avant [tanwi:n fatxa], l'Alif est omis. La règle s'écrit alors :

$$\text{« ا »} \longrightarrow [\Phi]/ \text{« ً »} _Y, \quad Y = \{\Phi, \text{graphème nul}\}.$$

c- Lorsque la "ا" (Alif) se produit au milieu comme étant hamzat Wasl, il est omis et la Hamzat Wasl se prononce comme /?/. La règle est sous la forme :

$$\text{« ا »} \longrightarrow [?]/ X_ \text{« ء »}, \quad X = \{\text{tout graphème}\}$$

Par exemple, le mot « طأطأ » (il hochait la tête). Les deux Alifs sont omis et le mot est prononcé comme /t'a?t'a?/.

d- Lorsque la « ا » (Alif), se produit au milieu du mot, il sert à vocaliser le graphème qui le précède immédiatement en ajoutant la voyelle longue /a:/. La règle prend la forme :

$$\text{« ا »} \longrightarrow [a:]/ \text{« فتحة »} _Y, \quad Y = \{\text{tout graphème représentant une consonne}\}$$

Par exemple, dans le mot « جواب » (réponse) le « ا » est précédé par " فتحة " et il apparaît après la "w" (« و »). Dans ce cas, le « ا » est utilisé pour vocaliser le "w" en générant la voyelle longue /a:/.

Le mot est prononcé comme /zawa:b/.

(3) **La règle du « alif maqsu:ra »** : Le remplacement du graphème « ي » qui apparaît en fin de mot, précédé par une « fatha », par la voyelle longue /a:/.

$$\text{fatha (« ُ ») + « ي »} \longrightarrow [a:]/ X_ \Phi, \quad X = \{\text{tout graphème représentant une consonne}\}$$

Par exemple, dans le mot « سلوى » /salwa:/ (Nom féminin), le graphème « ي » est à la fin du mot et est précédé par une voyelle courte « fatha ». Il est transformé en son de la voyelle longue /a:/.

(4) **Règle d'insertion de l'arrêt glottal (Glottalisation التهميز)**. Il existe trois sous-règles dans cette catégorie dont la plupart traitent de l'insertion de son occlusif glottal quand un mot commence ou se termine par « ء » (hamza), ou quand elle apparaît au milieu. Par exemple, les mots « أذهب » /ʔaDhab/ (je pars), « جاء » /ʔa:ʔa/ (il est venu), et « جاءت » /ʔa:ʔat/ (elle est venue) commençant, ou se terminant, ou ayant au milieu un hamza. Ces mots auront une occlusive glottale (Un coup glottal) accompagnant leurs réalisations phonétiques.

(5) **Règles [tanwi:n]**. En arabe standard, il y a certains graphèmes, dénommés «signes diacritiques [tanwi:n]» ou «symboles de vocalisation». Ils sont : [tanwi:n fatxa] « َ », [tanwi:nd'amma] « ِ », et [tanwi:n kasra] « ِ ». Ils apparaissent en fin du mot. Les prononciations régulières de ces [tanwi:n] sont, respectivement, /an/, /un/ et /in/, c'est à dire, correspondants à l'une des trois voyelles courtes /a/, /u/ ou /i/, suivie par le son nasal /n/. Dans cette catégorie, Il y a les trois règles :

« َ » → [an]/X_Φ, X = {tout graphème représentant une consonne}
 « ِ » → [un]/X_Φ, X = {tout graphème représentant une consonne}
 « ِ » → [in]/X_Φ, X = {tout graphème représentant une consonne}

(6) **Règle de gémation ou (Chedda)**. Chaque fois que la gémation diacritique « ّ » apparaît sur un caractère, ce caractère se dédouble en prononciation. Par exemple, dans le mot « جيد » (bon), le signe de germination apparaît au-dessus de « ي », ce caractère se répète donc, et le mot est prononcé comme /ʔajjid /. Les règles de « chedda » sont de la forme :

« ّ » → [X]/X_Y,
 X et Y = {tout graphème}.

(7) **Les règles de ligature en langue arabe**. Les ligatures en langue arabe sont des combinaisons de deux ou plusieurs symboles de caractères représentés en écriture par un symbole orthographique unique. Il y a sept sous-règles dans cette catégorie. Elles s'appliquent à "و" (hamza sur waw), "ي" (hamza sur yaa), "ا" (hamza sur alif), "ا" (alif mada), et le "لا" (hamza sur lam alif). Le « hamza sur waw » est prononcé comme étant un arrêt glottal suivi par la voyelle courte /u/. La « hamza sur yaa » est prononcée comme étant un arrêt glottal suivi par la voyelle courte /i/. La « hamza sur alif » est prononcé comme un arrêt glottal suivi par la voyelle courte /a/. Le « alif mada » est prononcé comme une voyelle longue /a:/, et le « Lam Alif Hamza » est prononcé comme /laʔa/. Les règles prennent la forme :

« و » → [ʔu]/X_Y, X,Y = {tout graphème représentant une consonne}
 « ي » → [ʔi]/X_Y, X,Y = {tout graphème représentant une consonne}
 « ا » → [ʔa]/X_Y, X,Y = {tout graphème représentant une consonne}
 « ا » → [ʔa:]/X_Y, X,Y = {tout graphème représentant une consonne}
 « لا » → [laʔa]/Φ_Y, Y = {tout graphème représentant une consonne}
 « لا » → [ʔalʔa]/«ا» (Alif)_Y, Y = {tout graphème représentant une consonne}
 « لا » → [ʔi]/X_Y, X,Y = {tout graphème représentant une consonne}

La ligature "لا" se trouvant dans certains claviers, peut être réalisée séparément à partir de "ل" (Lam) et "ا" (Alif), et généralement prononcé comme /la:/. Il faut noter qu'il y a des ligatures qui combinent la chedda et les symboles de voyelles (kasra, fatha ou damma) ou la chedda et les symboles du [tanwi:n]. Les prononciations de ce genre de ligatures se font par combinaison en série de la prononciation du chedda avec celle de la voyelle ou du [tanwi:n] respectivement.

(8) **Règles des Lettres solaires et des lettres lunaires** : Ces règles sont applicables aux lettres Chamsi (lettres solaires) et à leurs homologues Kamari (lettres lunaires), elles sont parmi les plus élaborées dans le système orthographique de la langue arabe standard. Les lettres solaires et lunaires ont été présentées dans la section 2.4.2 du chapitre précédent lors de l'examen des processus morpho-phonémiques. Ces règles portent sur la façon de traiter "ال" (Alif lam) lorsqu'il précède au début d'un mot soit un caractère solaire ou un caractère lunaire. Le «Alif lam» normalement prononcé comme étant /?al/, cette prononciation est altérée par la présence d'un caractère chamsi qui vient après. Si "Alif lam" se produit avant un caractère kamari, la prononciation normale par défaut, /?al/ est maintenu. Toutefois, s'il se produit avant un caractère chamsi il en va autrement, et la prononciation par défaut de l'«Alif lam» /?al/ est modifiée pour /?a/, et le caractère chamsi qui le suit devient géminé ou bien le « ل » (Lam) de «ال» est assimilé au son du caractère chamsi (genre d'assimilation «إدغام»). Cela dit, ce caractère chamsi géminé venant après un (Alif Lam) est regroupé de part sa propre nature pour former une sorte de groupe de consonnes géminées CC. En outre, si le mot précédent se termine par une voyelle, ce mot et le mot contenant le "ال" (Alif lam) sont fusionnés ensemble. Par exemple, le mot "الشمس" (le Soleil) quand prononcé isolément, se prononce comme /?aSSamsu/ tandis que le mot "القمر" (la lune) quand il est prononcé isolément, il se prononce comme /?alqamaru/. Toutefois, lorsque le mot "الشمس" est précédé par un mot qui se termine par une voyelle tels que l'expression « طلعت الشمس » (avec "كسرة" /i/ après le "ت" /t/), les deux mots sont fusionnés en une phrase composite, qui se prononce /t'ala?'atiSSamsu/. Étant donné que nous avons 28 caractères en langue arabe standard, il y aura 28 sous-règles dans cette catégorie de règles chamsi et kamari.

Les règles chamsi ont la forme :

«ال» → [?aYY]/Φ_Y, Y = {toute lettre solaire},

«ال» → [X?aYY]/X_Y, où Y = {toute lettre solaire} et X = {ف, و, د, ذ, ر, و, ف}

En effet, nous avons prévu tous les cas des lettres solaires préfixées par les différentes variantes du (Alif Lam) comme :

فـال، وال، بال، كال، وكال، وبال، فكال، فبال

Exemple : فَبِالنَّصْرِ se transcrit : /fabinnas'ri/

Puis entre deux mots :

«ال» → [XYY]/X_Y, où Y = {toute lettre solaire} et

X = {tout graphème de voyelle figurant à la fin du mot précédent},

Et les deux mots seront fusionnés (concaténés).

(Voir Exemple du paragraphe précédent : طلعت الشمس).

Les règles Kamari ont la forme :

«ال» → [?aℓ]/Φ_Y, Y = {toute lettre lunaire}.

Et si les lettres lunaires au début du mot sont préfixés par les différentes variantes du Alif lam (فـال، وال، بال، كال، وكال، وبال، فكال، فبال):

«ال» → [Xℓ]/X_Y, où X = {ف, و, د, ذ, ر, و, ف} et Y = {toute lettre lunaire}.

Exemple : Dans la phrase (فَبِالنَّصْرِ تَحِيًّا سَعِيدًا) le mot «الكَرَم» est préfixée par une variante du Alif Lam «فَبِالْ» et se transcrit par : /fabilkarami/.

Puis entre deux mots :

«ال» → [Xℓ]/X_Y, où

X = {tout graphème de voyelle figurant à la fin du mot précédent},

Y = {toute lettre lunaire}.

Et les deux mots seront fusionnés (concaténés).

Exemple : « خُسُوفُ الْقَمَرِ » dont la transcription est : /Xusufulqamari/. Nous rencontrons, parfois, une hamza placée sur l'alif de « alif lam » « ال », comme dans "أل". Ce hamza sur alif est une « /hamzat wasl/ » qui signifie une fusion entre le mot actuel et son prédécesseur.

(9) **Règles de génération de voyelles longues.** Il existe trois sous-règles dans cette catégorie :

- "ـَ" (fatha) avant "ا" génère la voyelle longue /a:/;
- "ـِ" (kasra) avant "ا" génère la voyelle longue /i:/
- et "ـُ" (damma) avant "ا" génère la voyelle longue /u:/.

Ces règles se présentent sous les formes :

- "ـَ" (fatha) \longrightarrow [a:]/X_Y,
X = {tout graphème représentant une consonne}, Y = {«ا»},
- "ـِ" (kasra) \longrightarrow [i:]/X_Y,
X = {tout graphème représentant une consonne}, Y = {«ي»},
- "ـُ" (damma) \longrightarrow [u:]/X_Y,
X = {tout graphème représentant une consonne}, Y = {«و»}.

(10) **Règles de génération de diphtongues.** Il existe deux sous-règles dans cette catégorie, une pour chacune des deux diphtongues (/aj/ et /aw/) de la langue arabe standard. Par exemple, la séquence graphémique « الفتحة » + « ي » dans "جَيِّد" /ʒajjid/ (bon). La rencontre du graphème de la syllabe fermée « ي » (correspondant au phonème /j/) avec fatha « فتحة » qui le précède est transformée en groupement sonore de la diphtongue /aj/. La règle est de la forme suivante :

"ي" « j » \longrightarrow [aj]/ « فتحة » _ « ي », lors de la fermeture d'une syllabe de type CVC.

Une règle similaire s'applique à la diphtongue /aw/. Par exemple, la séquence graphémique « الفتحة » + « و » dans "جَوَاد" /ʒawa:d/ (cheval). La rencontre du graphème de la syllabe fermée « و » "w" avec « فتحة » qui le précède est transformée en groupement sonore de la diphtongue [aw]. La règle est de la forme suivante :

"و" « w » \longrightarrow [aw]/ « فتحة » _ « و », lors de la fermeture d'une syllabe de type CVC.

Au cours de la mise en œuvre des règles (*Chap. 4*), l'application des règles des diphtongues est reportée jusqu'à ce que la syllabation soit effectuée. Ceci est important parce que la fin des syllabes fermées doit être détectée avant d'utiliser la règle des diphtongues pour détecter les diphtongues eux-mêmes.

(11) **Règles de remplacement des voyelles courtes.** Dans cette catégorie, les règles sont au nombre de trois, et concernent chacune le graphème d'une voyelle courte (/fatxa/, /d'amma/ et /kasra/). Chaque graphème d'une voyelle courte qui n'est pas encore consommé (c'est à dire qui n'est pas encore traité par une autre règle précédente) est remplacé par sa transcription sonore. Par exemple, le graphème "ـَ" (fatha) est remplacé par le phonème /a/. Le graphème "ـِ" (kasra) est remplacé par /i/, et "ـُ" (damma) est remplacé par /u/ :

- "ـَ" (fatxa) \longrightarrow [a]/X_Y,
X = {tout graphème représentant une consonne},
Y = {tout graphème représentant une consonne ; à l'exclusion du Alif «ا»},
- "ـِ" (kasra) \longrightarrow [i]/X_Y,
X = {tout graphème représentant une consonne},
Y = {tout graphème représentant une consonne ; à l'exclusion du « j » « ي »},
- "ـُ" (damma) \longrightarrow [u]/X_Y,
X = {tout graphème représentant une consonne},
Y = {tout graphème représentant une consonne ; à l'exclusion du « w » « و »},

Les exclusions mentionnées dans les contextes Y, ne sont pas indispensables si les règles de production de voyelles longues sont appliquées avant les règles des voyelles courtes. Si cela est le cas, tout fatha précédant "ا", damma précédant "و", ou kasra précédant "ي" sera consommée (traitée) par l'application des règles de génération des voyelles longues.

Nous présentons l'algorithme en pseudo-code de réalisation des règles Graphème-Phonème susmentionnées. Les commentaires qui s'y trouvent indiquent la position où certaines règles sont appliquées.

```
// Algorithme Graphème-Phonème
//Application des règles : Soukoun, d' élision, Ya Maqsourah, de ligature, [tanwi:n] et de gémination
// Les règles de 1 à 7
Lire la séquence du mot à prononcer dans une mémoire tampon graphèmes ;
Créer un tampon de mot phonologique de sortie vide ;
While (non fin de la mémoire tampon graphèmes)
{
    Lire le caractère suivant de la mémoire tampon ;
    // suppression du soukoun et élision
    If (le caractère satisfait la règle d' élision ou de suppression)
    {
        Supprimer la caractère de la mémoire tampon
    }
    // Ya Masourah
    Else if (char est le dernier caractère du mot ET le caractère est une Ya Maksourah)
    {
        Remplacer char par la voyelle /a/
    }
    //Ligature
    Else if (char est une ligature)
    {
        Insérer la séquence phonémique de ligature ;
    }
    // [tanwi:n]
    Else if (char est un caractère [tanwi:n])
    {
        Insérer la séquence phonémique appropriée pour [tanwi:n]
    }
    //Gémination
    Else if (char est le signe de gémination)
        Répéter et insérer le caractère avant le signe de gémination
    }
}
Continue
```

Pseudo-code de la Conversion Graphème-Phonème

```

// Algorithme Graphème-Phonème (...suite )
// Parcours de la mémoire tampon des graphèmes de gauche à droite
While (Il y a un mot dans le tampon des graphèmes) ;
{
    Mettre le tampon du mot phonémique de sortie à NULL ;

    // La recherche dans le lexique des mots exceptionnels
    If (le mot graphémique est dans le lexique des mots exceptionnels)
    {
        Lire l' orthographe du mot à partir du dictionnaire ;
        Concaténer le mot phonémique de sortie en utilisant son orthographe ;
    }

    // Application des règles restantes selon leurs ordre (règle de 8 à 11, Chamsi and Kamari,
    // les voyelles longues, diphtongue et la génération de voyelles courtes
    Else if (le mot graphémique n' est pas dans le lexique des mots exceptionnels)
    {
        Positionner la fenêtre focus au début du mot graphémique ;
        Initialiser la taille de la fenêtre focus à M (M est la plus grande taille dans le jeu de
règle) ;
        // Parcours du mot graphémique courant de gauche à droite par M, M-1, Jusqu' à 1
        While (M n' est pas inférieur à 1)
        {
            Positionner la fenêtre focus au début du mot graphémique ;
            Lire M caractères du mot graphémique ;
        }
        While (non fin du mot graphémique)
        {
            // Parcours la liste des règles ordonnées de haut en bas dans la manière dont elles sont
            //classées, les règles de grande taille en premier.
            While (non fin de la liste de règles)
            {
                If (Il y a une règle correspondante au M caractères dans la fenêtre focus)
                {
                    Lire les phonèmes correspondants à la règle ;
                    Concaténer le mot phonémique de sortie avec les phonèmes correspondant
                    Application de la règle ;
                    Sauter à droite dans le mot graphémique en lisant M caractères
                }
            }
            Décrémenter M par 1 ;
        } //fin des blocs else if (..)
    }
}
Appeler l' algorithme Syllabation_Diphtongue_Generation

```

Pseudo-code de la Conversion Graphème-Phonème ...Suite et fin

3.2.2 Règles phonétiques

Des règles supplémentaires phonème-phone sont utilisées pour couvrir les variantes phonétiques (allophoniques) essentielles des sons de la langue arabe. Les variations phonétiques importantes des sons arabes comprennent les phénomènes de coarticulation anticipatoires, et les phénomènes d'assimilations.

3.2.2.1 Coarticulations anticipatoires

Il y a coarticulation anticipatoire lorsque l'articulateur anticipe son déplacement vers le prochain lieu d'articulation, de rétention : le lieu d'articulation n'est pas atteint à cause de l'inertie de l'articulateur. Nous pouvons citer parmi ces phénomènes : la pharyngalisation de voyelles et diphtongues, la nasalisation de voyelles, et autres coarticulations anticipatoires comme le recouvrement des sons et les adaptations sonores [133, 134, 135].

- *Pharyngalisation*

La pharyngalisation des sons arabes est à la fois un phénomène hétéro-syllabique et tautosyllabique², il est également un phénomène progressif (en avant) ou régressif (en rétro). Elle affecte tous les six voyelles arabes (/a/, /i/, /u/, /a:/, /i:/ et /u:/)³, les deux diphtongues (/aw/ et /aj/), les phonèmes (/t/, /d/, /s/ et /D/) équivalents aux emphatiques de la langue arabe (/t'/, /d'/, /s'/ et / D'/, *figure 3.1*), la sonante / *l* / et le trille /r/ (vibrante alvéolaire). Les voyelles, les diphtongues, la sonante / *l* / et le trille / r / deviennent fortement pharyngalisés chaque fois qu'ils se produisent dans un contexte de pharyngalisation. Les phonèmes contreparties emphatiques (les opposées aux emphatiques) sont assimilées aux sons des emphatiques eux-mêmes. Des exemples de pharyngalisation de voyelles et de diphtongues, et d'assimilation emphatique de la langue arabe standard se trouvent dans les mots suivants :

- (1) "سطع" [s'a't'a'?'a] (brille). La première et la seconde voyelle courte /a/ sont pharyngalisées et prononcées comme [a']. Étant donné que la première voyelle courte /a/ est pharyngalisée, le phonème (س) /s/ est assimilé à son contrepartie (opposé) emphatique (ص) /s'/.
- (2) "سيطر" [s'aj't'a'ra] (a maîtrisé). La diphtongue /aj/ est pharyngalisée et se prononce comme [aj']. La deuxième voyelle courte /a/ est également pharyngalisée et se prononce comme [a'].
- (3) La pharyngalisation peut affecter certains sons avant le son pharyngalisé. Par exemple, dans le mot « مُسْتَطِيل » /mustat'i:l/ (rectangle), tous les sons /u/, /s/, /t/ et /a/ avant le /t'/ sont pharyngalisés.
- (4) Le son /d/ dans le mot « مِئْذَنَة » /mind'adah/ (table) est pharyngalisé même s'il n'apparaît pas dans la même syllabe qui contient le son pharyngalisé /d'/.

La pharyngalisation d'une consonne est la réalisation du son correspondant avec une articulation secondaire caractérisée par une constriction dans la paroi du pharynx produite par le recul de la racine de la langue. Cette pharyngalisation est accompagnée par une emphatisation. La pharyngalisation d'une voyelle est la réalisation de cette voyelle avec une constriction particulière dans le conduit vocal, *figure 3.2*. La voyelle [a] pharyngalisée par exemple est caractérisée par un F₁ très haut autour de 750 Hz et F₂ bas autour de 1300 Hz, F₁ et F₂ se trouvent donc rapprochés. La diminution de F₂ est causée par un élargissement de la cavité buccale, tandis que l'élévation de F₁ est due au rétrécissement dans la cavité pharyngale.

² - Hétérosyllabique signifie « n'appartenant pas à la même syllabe », et tautosyllabique signifie « appartenant à la même syllabe ».

³- Certaines études acoustiques ont montré que le système vocalique arabe traditionnellement décrit comme un ensemble de 6 voyelles (3 voyelles brèves et 3 voyelles longues correspondantes) est inadéquat. Un système de 12 voyelles dont 6 pharyngalisées est défini en compatibilité avec les faits acoustiques et phonologiques. Les règles régissant l'extension de la pharyngalisation au sein de la chaîne sonore sont établies et validées par synthèse (Voir [136]).

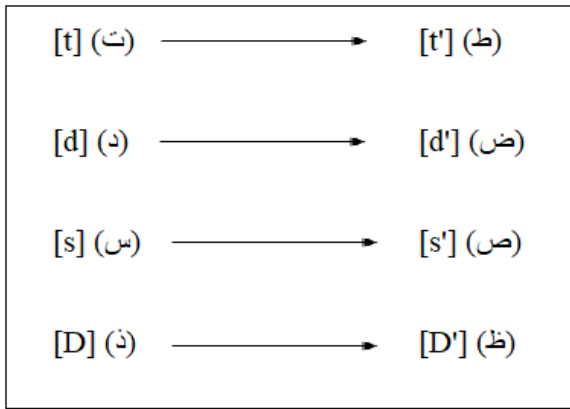


Fig. 3.1 : Paires des sons emphatiques et non emphatiques correspondants

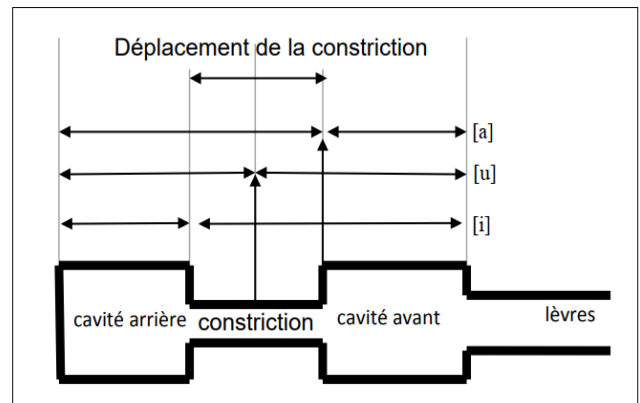


Fig. 3.2 : Modèle simplifié du conduit vocal avec déplacement de la constriction pendant la pharyngalisation d'une voyelle.

Les voyelles et les diphtongues de la langue arabe sont nasalisées quand elles rencontrent un son nasal (/m/ ou /n/). Les coarticulations anticipatoires comme le recouvrement sonore et l'assimilation auront lieu lorsqu'une consonne occlusive sourde /t/ ou /k/ est suivie par une voyelle longue fermée (postérieure ou palatale) /u:/ ou /i:/. Dans le contexte de la synthèse de la parole, ces variations phonétiques de la langue arabe sont amplement discutées dans [137].

Nous avons défini trois catégories comprenant vingt-deux règles phonème-phonème. Ces règles couvrent les altérations phonétiques les plus importantes décrites ci-dessus que subissent les phonèmes de base. Les règles de variations phonétiques sont :

Quatorze règles pour couvrir la pharyngalisation des six voyelles, des deux diphtongues, de la sonante /ℓ/ et du trille /r/ et l'assimilation des équivalents emphatiques. Ces sons subissent leurs changements phonétiques respectifs chaque fois qu'ils sont dans la même syllabe ou dans le voisinage d'une emphatique ou de tout autre « son » fortement pharyngalisé. Les règles de pharyngalisation des six voyelles sont de la forme :

[V] → [V']/X_Y où X, Y : Sont une syllabe immédiatement voisine contenant un son emphatique (« ط », « ظ », « ص », « ض » ou tout autre son hautement pharyngalisé) ou tout simplement X et Y sont un son pharyngalisé immédiatement voisin. V est l'un des six phonèmes vocaliques (/a/, /u/, /i/, /a:/, /u:/ et /i:/) et V' est l'équivalent pharyngalisé de V. Dans le cas de la coarticulation « Consonne + voyelle », les consonnes pharyngalisées (/t'/ ط, /d'/ ض, /s'/ ص et /D'/ ظ) exercent une agression coarticulatoire forte sur les voyelles adjacentes, comparées à leurs correspondantes non pharyngalisées (/t/ ت, /d/ د, /s/ س et /D/ ذ) qui coarticulent mieux avec les voyelles. Nous constatons la nécessité de la syllabation avant l'application de toute règle de pharyngalisation. Les voyelles pharyngalisées [i' a' u' i:' a:' u:] équivalent à leurs contreparties non-pharyngalisées [i a u i: a: u:], dérivent donc par assimilation d'une consonne avoisinante pharyngalisée.

Pour les diphtongues, les règles sont de la forme :

[aj ou aw] → [aj' ou aw']/X_Y; X,Y= {même contexte donné aux cas des voyelles ci-dessus}

aj ou aw sont des diphtongues non-pharyngalisées et aj' ou aw' sont leurs homologues pharyngalisées. Un exemple de pharyngalisation de la diphtongue /aj/ a été déjà donné plus haut. Nous pouvons observer la pharyngalisation de la diphtongue /aw/ dans le mot : " طَوْر " [t'aw'war] (a développé). La diphtongue /aw/ est pharyngalisée et est prononcée [aw'] en raison de la présence de l'emphatique /t'/ dans la même syllabe qu'elle.

Pour l'assimilation des équivalents des sons emphatiques et la pharyngalisation de la sonante /ℓ/ ou du trille /r/, nous utilisons des règles de la forme :

[NE ou C] → [E ou C']/X_Y; X,Y= {même contexte donné aux cas des voyelles ci-dessus} NE est tout « son » opposé au « son » emphatique équivalent et E est le « son » emphatique lui-même. C est soit la sonante /ℓ/ ou le trille /r/ et « C' » est son équivalent pharyngalisé.

- *Nasalisation des voyelles*

Six règles se chargent de la nasalisation des six voyelles. Ce phénomène apparaît chaque fois que l'un de ces sons est suivi par un son nasal (/m/ ou /n/). Ces règles s'écrivent sous la forme :

[V] → [Vⁿ]/X_Y, X= {tout graphème représentant une consonne}, Y = {m ou n}.

V est l'un des six phonèmes vocaliques (/a/, /u/, /i/, /a:/, /u:/ et /i:/). Vⁿ est la contrepartie nasalisée de V.

Ce phénomène de nasalisation est dû au fait que le voile du palais nécessite un certain temps pour s'abaisser afin de préparer la production de la consonne nasale cible (/m/ ou /n/), et partiellement pour se relever. Par conséquent une voyelle environnant une consonne nasale pourrait partiellement se contaminer par le trait nasal de la consonne. Il est à noter que cette contamination est plus ou moins forte, d'où l'on parle de « degré de nasalisation ». Il y a des altérations dans la configuration du conduit vocal lors du passage de la voyelle à l'articulation de la consonne nasale. Acoustiquement, les altérations d'une voyelle orale à une voyelle nasalisée se manifestent par :

- Une diminution des amplitudes du deuxième formant et des formants supérieurs au cours des 30-40 ms avant le début de la fermeture nasale alvéolaire ou vélaire ;
- Un étalement spectral notable dans la région des formants bas (notamment dans F1 et F2) en raison de la nasalisation des voyelles ;
- et un décalage dans les valeurs des formants des voyelles et l'apparition possible de pôles et de zéros supplémentaires dans la fonction de transfert qui représente le conduit vocal combiné.

Nous avons les exemples suivant de nasalisation des voyelles :

- Pour la voyelle [a] ; [Xaⁿnq] (étranglement), [ʔaⁿnfusahum] (eux-même) et [Xaⁿmsi:n] (cinquante) ;
- Pour la voyelle [u] ; [kuⁿntum] (vous étiez), [s'uⁿmtu] (j'ai jeûné) ;
- Pour la voyelle [i] ; [biⁿnt] (fille), [tiⁿmsa:hun] (crocodile) ;
- Pour la voyelle [a:] ; [fursa:ⁿnun] (des cavaliers) ;
- Pour la voyelle /u:/ : «ديون» [duju:ⁿnun] (dettes), «معلوم» [ma?'lu:ⁿmun] (déterminé)
- Pour la voyelle /i:/ : «عجين» [?'aʒi:ⁿn] (pâte), «حكيم» [xaki:ⁿm] (sage)

La Pharyngalisation est une propriété phonétique dominante de l'AS, de telle sorte que si elle affecte un son en même temps qu'un autre phénomène, la pharyngalisation dominera. Par exemple, dans le mot "صَنَّع" /s'ana?'a/ (a fabriqué), la première voyelle /a/ est à la fois pharyngalisée et nasalisée. En raison de la forte influence de la pharyngalisation, la voyelle est considérée comme pharyngalisée. Pour cette raison, les règles phonétiques sont également agencées (ordonnées) de manière à ce que les règles de pharyngalisation soient exécutées en dernier. Ceci et d'autres considérations ont des implications sur le type de structure de données utilisée pour représenter la chaîne phonétique. L'algorithme de réalisation de ces règles phonétiques est présenté en pseudo-code indiqué sur l'encadré suivant :

```

// Algorithme de conversion Phonème-Allophone
// Application des règles de nasalisation
Lire la liste doublement chaînée de la syllabation et de la génération de diphtongues;
Commencez par le nœud le plus à gauche ou le plus à droite
Repeat
  Recherchez dans la liste de gauche à droite ou de droite à gauche;
  Récupérer le nœud suivant de la liste;
  If (le contenu du nœud est nasal (/m/ ou /n/ )
  {
    Lire le nœud précédent;
    If (le nœud précédent est une voyelle)
    {
      Supprimer le nœud;
      Insérer le nœud dont l'élément de données est le symbole phonétique
de la voyelle nasalisée ;
    }
  }
  Jusqu'à fin de la liste
// Algorithme de conversion Phonème-Allophone (...suite)
// Application des règles de la coarticulation anticipatoire
Lire la liste doublement chaînée du processus nasalisation;
Placez-vous sur le nœud le plus à gauche ou le plus à droite;
Repeat
  Recherchez dans la liste de gauche à droite ou de droite à gauche;
  Récupérer le nœud suivant de la liste;
  If (le contenu du nœud est une occlusive sourde /t/)
  {
    Lire le nœud suivant;
    If ( le nœud suivant est une voyelle longue /u:/)
    {
      Supprimer le nœud précédent ;
      Insérer le nœud dont l'élément de données est le symbole
phonétique de la coarticulation / t /;
    }
  }
  If (le contenu du nœud est une occlusive sourde /k/)
  {
    Lire le nœud suivant;
    If ( le nœud suivant est une voyelle longue /i:/)
    {
      Supprimer le nœud précédent ;
      Insérer le nœud dont l'élément de données est le
symbole phonétique de la coarticulation / k /;
    }
  }
  Jusqu'à fin de la liste
Continue

```

```

// Algorithme de conversion Phonème-Allophone (...suite)
// Application des règles de Pharangalisation
Lire la liste doublement chaînée du processus de coarticulation anticipatoire;
Commencez par le nœud le plus à gauche ou le plus à droite
Repeat
  Recherchez dans la liste de gauche à droite ou de droite à gauche;
  Récupérer une syllabe de la liste en récupérant l'ensemble de nœuds entre les
balises (étiquettes) de syllabes;
  If (la syllabe contient un emphatique )
  {
    Remplacer le noyau de la syllabe par une voyelle pharyngalisée ;
    If (la syllabe contient /l/ ou /r/ )
    {
      Supprimer /l/ ou /r/ ;
      Remplacer /l/ ou /r/ par leurs contrepartis pharangalisés ;
    }
    If (La syllabe contient une contrepartie emphatique )
    {

```

```

        Supprimer la contrepartie emphatique
        Remplacer la contrepartie emphatique par le son emphatique
    }
}
Until fin de la liste.
If (Application requiert une chaîne phonétique syllabique non marquée (balisée))
    Supprimer les marques (balises) syllabiques de la liste ;
Transmettre la liste phonétique à l'application
Fin ;

```

Pseudo-code de la Conversion Phonème-Phone

3.2.2.2 Assimilations ou Règles du « [nu:n?asa:kina] »

La langue AS est une langue très infléchi. Les Inflexions ont un rôle phonologique essentiel dans la langue en général et dans la récitation des versets du Coran, en particulier.

Nous allons décrire certaines des règles phonologiques régissant l'articulation du son « *noune* » finale silencieuse (non vocalisé) de la lettre [nu:n?asa:kina] « نْ », et le "[tanwi:n]" inflexionnel. Nous avons déjà traité le cas intéressant du [tanwi:n] (Chapitre 2, Section 2.4.1), mais dans cette section, il s'agit aussi d'étudier l'influence du [nu:n?asa:kina] final et du [tanwi:n] sur le mot suivant, d'où l'apparition de plusieurs phénomènes phonologiques de frontières entre les mots à prendre en considération dans le système de transcription. Ces règles se résument dans la figure 3.3 ci-dessous.

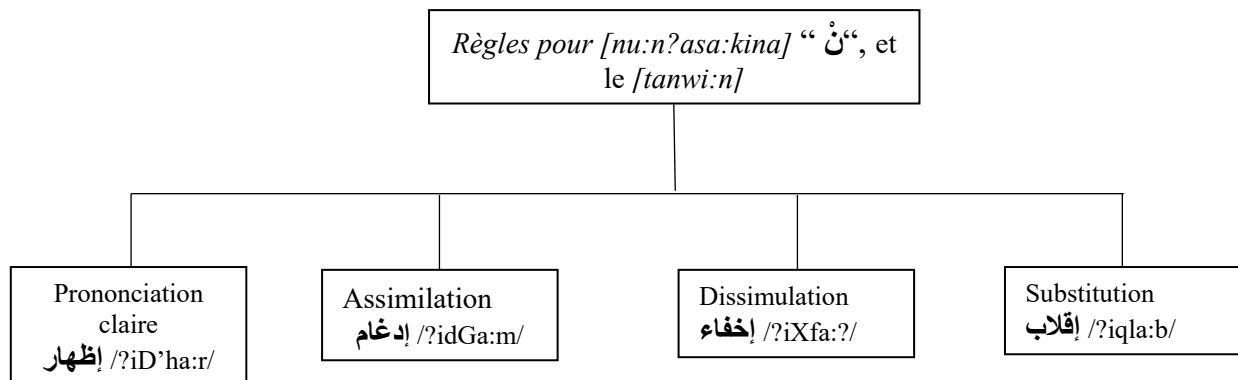


Figure 3.3 : Règles du [nu:n?asa:kina] et du [tanwi:n]

En effet, du point de vue phonologique, un [nu:n?asa:kina] – « نْ » et un [tanwi:n] sont considérés comme étant les mêmes car au niveau de leur sonorité, le [tanwi:n] sonne comme un [nu:n?asa:kina]. Ainsi, ils répondront aux mêmes règles de prononciation (Tableau 3.1).

- *Assimilation* : est définie par un grand nombre de phonéticiens comme Shane dans [138], Ladefoged dans [139] et beaucoup d'autres en tant que processus phonologique où il s'opère un changement d'un son en un autre à cause du rapprochement phonétique et du voisinage articulaire de ces deux sons. « Une consonne peut contracter les caractéristiques d'une voyelle, une voyelle peut prendre les caractéristiques d'une consonne, une consonne peut influencer une autre, ou une voyelle peut avoir un effet sur une autre » (Shane [138], p.49). Cette définition est confirmée par Fromkin dans [140] en notant que « l'assimilation est une sorte d'économie dans le processus d'articulation, dans lequel un son influence la prononciation d'un autre son adjacent ou à proximité » (p. 481). L'assimilation phonique est donc un type très fréquent de variation phonologique subie par un son au contact d'un son voisin et qui tend à réduire les différences entre ces deux sons : l'un des deux sons acquiert une ou plusieurs caractéristiques propres au son voisin. À l'inverse, lorsque deux sons semblables, en contact plus ou moins direct, s'éloignent l'un de l'autre, on parle de la dissimulation.

- *Le [nu:n?asa:kina] « نْ »* : Le "noûn sâkina" : النون الساكنة est le noûn dépourvu de voyelle brève. Sa prononciation est influencée par la lettre qui le suit.

- [tanwi:n] ou la nunation « التتوين » : c'est un noûn sâkina ajouté à la fin d'un mot dans la prononciation mais pas par écrit. Il n'est pas prononcé si on s'arrête sur le mot sur lequel il est situé, mais il l'est dans le cas contraire (lorsqu'on ne fait pas d'arrêt). Son symbole est le doublement de voyelle brève (deux /d'amma/ [un], deux /fatxa/ [an] ou deux /kasra/ [in]).

Tableau 3.1 : Réalisation phonétique du [tanwi:n]

Notation orthographique du [tanwi:n]	Articulation de cette Notation	Exemple	Transcription phonétique en SAMPA
ـَ	an	عزيرًا	[?'azi:zan]
ـُ	un	عزيرًا	[?'azi:zun]
ـِ	in	عزيرًا	[?'azi:zin]

- *Prononciation claire* ([?al?iD'ha:r]) الإظهار : Il consiste à sortir chaque son de son lieu d'articulation sans avoir recours à la nasalisation. Le lieu d'articulation (/maXra3/) du son /n/ est le bout de la langue contre le palais proche des incisives supérieures. Cette règle concerne 6 phonèmes appelés « les sons du larynx » (/xalqija/) :

هـ ع ح غ خ /h/, /?/, /?'/, /x/, /G/, /X/

Tableau 3.2 : Exemples de [?al?iD'ha:r]

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
مِنْ إِلَهٍ	مِنْ إِلَهٍ	min ?ila:hin
مِنْ خَوْفٍ	مِنْ خَوْفٍ	min Xawfin

- [?al?iqla:b] (*Substitution*) الإقلاب : Selon les phonéticiens Aljamal dans [141] et Nassir dans [142], [?al?iqla:b] correspond au changement du “son” de son point de sortie initial, ainsi, le [nu:n?asa:kina] – ن ou le [tanwi:n] se change en Mîm [m] – م. On appliquera cette règle lorsqu'on est en présence d'un [nu:n?asa:kina] – ن ou d'un [tanwi:n], et il est suivi par le phonème [b] – ب. Le résultat de cette rencontre est le changement du [nu:n?asa:kina] ou le [tanwi:n] en Mîm [m], avec une nasalisation-غنة de deux temps. Nous notons cette réalisation /mⁿ/. Ce signe diacritique (n) que nous avons adopté dans cette étude sera utilisé pour représenter ce phénomène de nasalisation (الغنة) qui accompagne cette substitution dans la représentation de la prononciation.

Tableau 3.3 : Exemples de « [?al?iqla:b] »

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
مَشَاءٍ بِنَمِيمٍ	مَشَاءٍ ⁿ بِنَمِيمٍ	maSa:?'im ⁿ binami:m
أَنْ بَوْرِكَ	أَمْ ⁿ بَوْرِكَ	?am ⁿ bu:rika
سَمِيعٌ بَصِيرٍ	سَمِيعِ ⁿ بَصِيرٍ	sami:?'um ⁿ basi:r

Nous avons un changement dans le lieu d'articulation. Au lieu de la position de la crête alvéolaire du son /n/, nous avons une position presque bilabiale, où les lèvres ne sont plus fermées hermétiquement comme pour le son /m/, ni complètement écartées comme pour le son nasal /n/, mais elles prennent une position médiane qui pourrait être définie comme semi-bilabiale /mⁿ/ permettant à une partie de l'air de passer à travers la cavité nasale et une autre partie à passer à travers la voie orale puisqu'il n'y a pas fermeture complète.

Le nouveau son /mⁿ/ ne se produit pas clairement parce qu'une autre règle dite « [ʔalʔiXfa:ʔ] » (Dissimulation) qui sera discutée plus tard, sera applicable ici. En résumé, cette règle pourrait être énoncée comme suit :

+ Nasal	<i>suivi par une occlusive</i>	+nasal
+ Alvéolaire	<i>bilabiale</i>	+ semi-bilabial
+ syllabe finale	→	

Il faut noter que ce phénomène apparaît dans les frontières entre les mots, comme il peut apparaître à l'intérieur du mot comme dans « **عنبر** » /ʔ'anbar/ qui se prononce « **عمبر** » /ʔ'amⁿbar/. En d'autres termes, et selon notre formalisme de codification, nous avons la règle suivante :

$$[n] \longrightarrow [m^n] / X_ [b]$$

Avec X={tout graphème}

- [ʔidGa:m] (Gémination : une forme d'assimilation) الإدغام : Quand l'assimilation du [nu:nʔasa:kina] – ن ou du [tanwi:n] s'effectue en présence des sons suivants :

ن - و - ل - م - ر - ي /j/, /r/, /m/, /l/, /w/, et /n/.

Ce phénomène de [ʔidGa:m] en langue arabe désigne la suppression du son nasal alvéolaire du [nu:nʔasa:kina] – ن ou du [tanwi:n] à chaque fois qu'il se produit dans la position finale d'un mot et est suivi par un mot commençant par un de ces six sons précités. En conséquence, ces sons seront gémérés **مشددة**. Dans l'écriture, la gémération est représentée par le doublement de la lettre gémérée. Mais il est intéressant de noter qu'aucune assimilation n'aura lieu si ces sons se produisent au sein d'un mot.

Il existe deux types de cette forme d'assimilation : [ʔidGa:m] sans nasalisation **الإدغام بغير غنة** et [ʔidGa:m] avec nasalisation – **الإدغام بغنة**

- [ʔidGa:m] avec nasalisation : Il est également appelé l'assimilation incomplète. Les sons de l'[ʔidGa:m] avec nasalisation sont au nombre de quatre ; elles sont regroupées dans le mot : ن - م - و (/j/, /n/, /m/ et /w/). L'assimilation avec nasalisation signifie que le son du [nu:nʔasa:kina] – ن ou du [tanwi:n] sera entendue de manière incomplète ou partiellement assimilée. La nasalisation reste comme un indicateur de la suppression du /n/. C'est une assimilation régressive mais non progressive qui se passe ici car il y a une absorption complète d'un phonème final dans un autre phonème initial [143].

Tableau 3.4 : Exemples du [ʔidGa:m] avec nasalisation

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
مَنْ نَعْمَةٍ	مَنْعَمَةٌ	min ⁿ niʔ'matin
مَنْ يَعْمَلُ	مَنْعَمَلٌ	maj ⁿ jaʔ'mal
مِنْ وَالٍ	مِنْوَالٌ	miw ⁿ wa:l
صِرَاطٌ مُسْتَقِيمٌ	صِرَاطٌ مُسْتَقِيمٌ	s'ira:tim ⁿ mustaqi:min
أُمَّةٌ وَاحِدَةٌ	أُمَّتٌ وَاحِدَةٌ	ummataw ⁿ wa:xida
قَبْضًا يَسِيرًا	قَبْضِيَّ سِيرًا	qabdaj ⁿ jasi:ra

Cette règle peut s'écrire sous forme de :

+ Nasal + alvéolaire + position finale	Suivi d'un bilabial/ <u>alvéolaire nasal</u> → ou (bilabial/ palatal) glide	- articulation du son final du mot précédent + nasalisation du premier son suivant + Gémination du premier son suivant
--	---	---

- [ʔidGa:m] sans nasalisation : [ʔidGa:m] sans nasalisation se fait lorsque la lettre [nu:nʔasa:kina] – نْ ou le [tanwi:n] à la fin d'un mot sont suivis par les lettres Ra – ر /r/ ou Lam – ل /l/ d'un autre mot (Les deux liquides: la latérale [l] et la vibrante[r]). La prononciation se fera sans nasalisation. En écriture, les lettres arabes Lam – ل et Ra – ر porteront alors le signe diacritique de gémination “*la chedda* ّ”.

On parle donc d'une assimilation complète car on ne retrouvera aucune trace sonore de la lettre Noune – ن. Nous parlons ici également, d'une assimilation régressive et non progressive car il y a fusion complète du son final dans le son initial avec annulation de tout trait phonétique du son supprimé.

Tableau 3.5 : Exemples du [ʔidGa:m] sans nasalisation

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
مَنْ رَبِّكَ	مَنْرَبِّكَ	mirrabika
أَنْ لَوْ	أَلَّوْ	?allaw
وَيْلٌ لِّكُلِّ	وَيْلَلِكُلِّ	wajlullikulli
غَفُورًا رَحِيمًا	غَفُورَرَّحِيمًا	Gafu:rarraxi:ma

La règle ci-dessus peut être énoncée comme suit :

+ Liquide + Position initiale	précédé par le son final /n/ →	+ gémination - Nasalisation - le /n/ final
----------------------------------	--------------------------------	--

Cependant, certains chercheurs et phonéticiens arabes classent le processus d'assimilation connu comme « [ʔidGa:m] » en deux types différents : complet et incomplet, et ceci en fonction du degré d'assimilation.

- [ʔidGa:m] *total* الإِدْغَامُ الكَامِلُ est, en règle générale, l'insertion d'un son non vocalisé – c'est à dire non-infléchi, dans un son vocalisé voisin tel que le premier son disparaît complètement avec tous ces traits caractéristiques, avec gémation du deuxième. Dans la prononciation, cela revient à géminer les sons /l/ et /r/ quand ils viennent d'abord après un son final alvéolaire nasal /n/. Ici, ni nasalisation n'est produite, ni aucun autre trait de /n/ persiste.

Ceci est également valable pour les deux sons /m/ et /n/ quand ils apparaissent au début des mots précédés par un mot avec le son nasal final /n/. Il en résulte la gémation de ces deux sons /m/ et /n/ avec une nasalisation, ce qui est très évident. Cependant, cette nasalisation n'est pas un trait acquis de la suppression du /n/ mais provient des sons initiaux /m/ et /n/. Les exemples suivants illustrent ce phénomène.

Tableau 3.6 : Exemples du [ʔidGa:m] total

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
مِنْ لِدْنِه	مِلْدَنِه	miɫɫadunh
مِنْ رِبِّهِمْ	مِرْرَبِّهِمْ	mirrabihim
كَأْسٍ مِنْ مَعِينٍ	كَأْسِمَمَّعِينٍ	kaʔsim ⁿ mmaʔ'i:n
مَنْ نَصِيرٍ	مِنِّصِيرٍ	min ⁿ nasi:rin
مَالًا لِبَدَا	مَالِلْبَدَا	ma:lalɫubada
فِي عَيْشَةٍ رَاضِيَةٍ	فِي عَيْشَتَّرَاضِيَةٍ	fi:ʔ'ɪ:Satirradijah
مَنْ مَاءٍ مِهِينٍ	مَمَائِمَّهِينٍ	mim ⁿ ma:ʔim ⁿ mahi:n
صَابِرًا نَعَمَ الْعَبْدِ	صَابِرَتَّعَمَ الْعَبْدِ	s'a:biran ⁿ niʔ'malʔ'abdu

- [ʔidGa:m] *Partiel (Assimilation 2)* : C'est la fusion de deux sons voisins (dans les frontières entre deux mots distincts) qui se traduit par la suppression du premier son et la préservation du second en maintenant la nasalisation en tant que marque du son supprimé [141]. Dans la prononciation des textes en AS, ce phénomène apparaît quand les sons /w/ et /j/ ("و" et "ي") se produisent au début des mots succédant un mot avec le son final alvéolaire nasal /n/. Prenons les exemples suivants :

Tableau 3.7 : Exemples du [ʔidGa:m] partiel

Transcription Orthographique	Prononciation	Transcription phonétique en SAMPA
وَمَنْ يُؤْمِنُ	وَمِيؤْمِنُ	wa maj ⁿ ju:min
مِنْ وِرَاءِ	مَوْرَاءِ	miw ⁿ wara:ʔ
خَيْرًا يَرَهُ	خَيْرِيَرَهُ	Xajraj ⁿ jarah
مَالٍ وَبَنِينَ	مَالَوْبَنِينَ	ma:liw ⁿ wa bani:n

- [ʔalʔiXfa:ʔ] الإخْفَاءُ (Assimilation 3 : La dissimulation): La langue AS possède 15 sons connus comme étant des « sons faibles » parce qu'ils sont traités différemment lorsqu'ils sont précédés d'un mot se terminant par le son alvéolaire nasal non vocalisé /n/. Ces sons comprennent tous les occlusifs sauf /b/ et /ʔ/, les fricatifs, le son occlusif-fricatif /ʒ/ et les emphatiques, c'est-à-dire les 15 sons qui ne sont ni « al-halqija », ni /b/ ni les sons de la parole « yarmaloun /j/,/r/,/m/,/l/,/w/,/n/ » [144].

Tableau 3.8 : Les 15 « sons faibles » de la l'AS.

Occlusifs	/t/ = (ت), /d/ (د), /k/ (ك)
Fricatifs	/s/ (س), /z/ (ز), /f/ (ف), /T/ (ث), /D/ (ذ), /S/ (ش)
Occlusif-fricatif	/ʒ/ (ج)
Empathiques	/s'/ (ص), /d'/ (ض), /t'/ (ط), /D'/ (ظ), /q/ (ق)

Chaque fois que le son nasal alvéolaire /n/ du “*n non vocalisé*” ou [tanwi:n] est suivi par l'un des 15 sons mentionnés ci-dessus, un processus d'assimilation aura lieu. Dans un tel cas, le /n/ va changer son point d'articulation, mais conserve son trait de nasalisation. Pour le dire autrement, le bout de la langue ne touchera pas la crête alvéolaire supérieure, mais il se déplace plus près du lieu d'articulation du son suivant. Ensuite, il restera dans une position neutre en face de ce lieu un moment suffisant pour produire la nasalisation qui donne à l'auditeur une impression implicite du son /n/. Dans ce cas, /n/ est maintenu, non assimilé, mais il n'est pas explicitement prononcé. La langue revient immédiatement dans sa nouvelle position articuloire. Nous sommes devant un phénomène d'assimilation régressive partielle.

Le phénomène de dissimulation « [ʔalʔiXfa:ʔ] » الإخفاء apparaît soit à l'intérieur du mot ou dans les frontières entre les mots chaque fois que le son /n/ non vocalisé ([nu:nʔasa:kina]) est suivi par l'un des 15 sons mentionnés ci-dessus, alors que dans le cas du [tanwi:n], il n'aura lieu qu'entre les mots mais pas à l'intérieur du mot. En outre, il est intéressant de noter que lorsque le son /n/ est suivi par l'un des sons emphatiques : /t'/, /d'/, /s'/, /D'/ (en plus de /q/ pour certains auteurs), une modification supplémentaire ait lieu. Dans ce cas de figure, la pointe et l'avant de la langue restent dans une position à mi-chemin en face du lieu d'articulation du son suivant ; quant à l'arrière de la langue tend à toucher le palais supérieur. Donc, l'air ne peut pas passer librement parce qu'il rencontre une langue soulevée. En conséquence, un son emphatique est produit. Nous avons donc la règle formelle suivante :

[n] —————> [nʰ]/X_[Y] ; où X={tout graphème} et [Y] est un son du Tableau 3.8.

Avec [nʰ] désignant le phénomène de « [ʔalʔiXfa:ʔ] » du [nu:nʔasa:kina]. Considérons les exemples suivants :

Tableau 3.9 : Exemples de [ʔalʔiXfa:ʔ] en [nu:nʔasa:kina]

Le son de [ʔalʔiXfa:ʔ]		Le son dissimulé [nu:nʔasa:kina] à l'intérieur du mot		[nu:nʔasa:kina] à la frontière entre les mots	
Lettre	Phonème	Mot	Prononciation (en transcription SAMPA)	Succession de deux Mots	Prononciation (en transcription SAMPA)
ت	t	كنتم	kun ^o tum	مِنْ تَسْنِيمٍ	min ^o tasni:m
ث	T	منثورا	man ^o Tu:ran	فَمَنْ ثَقَلَتْ	faman ^o Taqulat
ج	ʒ	فأنجيناها	faʔan ^o ʒajna:hu	مَنْ جَاءَ	man ^o ʒa:ʔ
د	d	أندادا	ʔan ^o da:dan	مِنْ دَابَّةٍ	min ^o da:batin
ذ	D	لتنذر	litun ^o Dira	مَنْ ذَا	man ^o Da:
ز	z	أنزل	ʔan ^o zala	مَنْ زَكَّاهَا	man ^o zakka:ha
س	s	الإنسان	ʔalʔin ^o sa:n	مِنْ سَوْءٍ	min ^o su:ʔin
ش	S	انشقت	ʔin ^o Saqat	فَمَنْ شَاءَ	faman ^o Sa:ʔa
ص	s'	ينصرون	jan ^o s'uru:n	مِنْ صَلَاحٍ	min ^o s' ils'a:lin
ض	d'	منضود	man ^o d'u:d	مِنْ ضَرِيحٍ	min ^o d'ari:ʔ'
ط	t'	ينطقون	jan ^o t'iqu:n	مِنْ طِينٍ	min ^o t' i:nin
ظ	D'	فلينظر	faljan ^o D'uri	مَنْ ظَلَمَ	man ^o D'alama
ف	f	ينفع	jan ^o fa:ʔ'	مِنْ فَضْلٍ	min ^o fad'li
ق	q	تتقمون	tan ^o qumu:n	مِنْ قَبْلِ	min ^o qabl
ك	k	منكم	min ^o kum	مَنْ كَفَرَ	man ^o kafara

Tableau 3.10 : Exemples de [ʔalʔiXfa:ʔ] (Assimilation 3) dans le cas du [tanwi:n]

Le son de [ʔalʔiXfa:ʔ]		[tanwi:n]	
Lettre	Phonème	Exemple	Prononciation (en transcription SAMPA)
ت	t	جناتٍ تجري	ʒanna:tin ^o taʒri:
ث	T	سحاباً ثقالاً	saxa:ban ^o Tiqa:la
ج	ʒ	عينٌ جارية	ʔʔajnun ^o ʒa:rijah
د	d	قنوانٌ دانية	qinwa:nun ^o da:nijah
ذ	D	يومٍ ذي	jawmin ^o Di:
ز	z	فاكهةٍ زوجان	fa:kihatin ^o zawʒa:n
س	s	بقلبٍ سليم	bicalbin ^o sali:m
ش	S	لنفسٍ شيئا	linafsin ^o Sajʔa
ص	sʔ	ريحاً صرصرا	ri:xan ^o sʔarsʔara
ض	dʔ	مسجداً ضرارا	masʒidan ^o dʔira:ra
ط	tʔ	سماواتٍ طباقا	sama:wa:tin ^o tʔiba:qa
ظ	Dʔ	ظلاً ظليلاً	Dʔillan ^o Dʔali:la
ف	f	لقولٍ فصل	laqawlun ^o fasʔl
ق	q	عذاباً قريباً	ʔʔaDa:ban ^o qari:ba
ك	k	رزقٍ كريم	rizqin ^o kari:m

En conclusion, nous pouvons dire que l'assimilation est un phénomène phonologique universel s'appliquant systématiquement dans la lecture des textes en langue arabe en général, et dans la récitation du Saint Coran en particulier. Il est réalisé sous trois formes : [ʔalʔiqla:b], [ʔalʔidGa:m], et [ʔalʔiXfa:ʔ]. Chacune de ces formes a son propre environnement d'occurrence et les conditions d'articulations. Le [ʔidGa:m] est une assimilation complète régressive telle que définie par Roach [145]. [ʔalʔiXfa:ʔ] est une assimilation régressive partielle, tandis que [ʔalʔiqla:b] est une coarticulation anticipatoire telle que définie par RAHIM et al. dans [146].

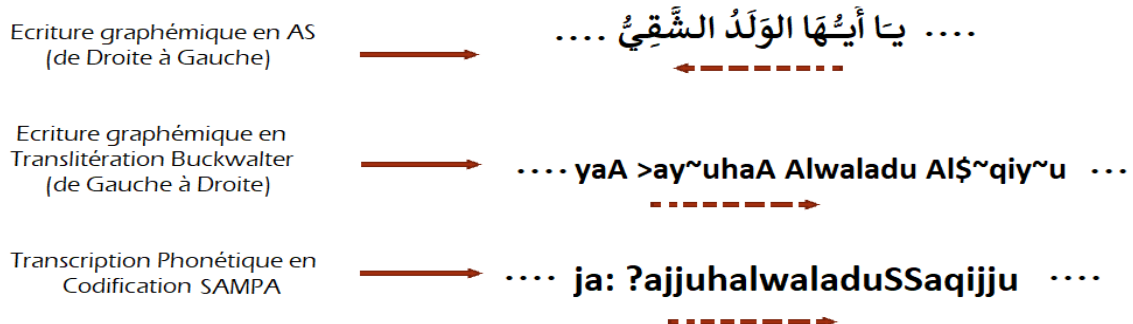
3.2.3 Réalisation des règles phonologiques et phonétiques

Les règles de pharyngalisation de certains sons arabes et celles de génération de diphtongues exigent que la chaîne phonémique soit organisée en syllabes. La Syllabation est la première étape à effectuer sur la séquence phonémique. Comme cela a été suggéré dans la *section 2.4.5* du chapitre 2, la séquence phonémique peut être segmentée en syllabes en notant que, en langue AS, le noyau de chaque syllabe est une voyelle. L'algorithme que nous proposons s'applique à la génération syllabique. La représentation phonémique est analysée à partir de son extrémité droite (avec un balayage de droite-à gauche) à la recherche de voyelles. Chaque fois qu'une voyelle est détectée, une vérification est déclenchée d'une correspondance syllabique pour l'un des six types de syllabes arabes (comme nous l'avons déjà mentionné, les syllabes arabes sont de type [CV], [CVC], [CV:], [CV:C], [CVCC] et le cas rare: [CV:CC], où C représente une consonne et V une voyelle). Chaque fois qu'une correspondance de syllabe est détectée, des balises (étiquettes) sont insérées pour marquer les frontières de syllabes. Cette étape est répétée jusqu'à ce que toutes les voyelles dans la séquence phonémique soient consommées. L'algorithme de la syllabation est présenté en pseudo-code représenté sur l'encadré plus bas. Le même code indique comment les diphtongues sont générées à partir des syllabes (si elles ne font pas partie de l'ensemble ordonné de règles phonémiques).

A titre d'exemple de la syllabation, considérons le mot CV.CV.CV.CVC "كُتُبُهُمْ" (leurs livres) dont la séquence phonémique est /kutubuhum/. L'application de l'algorithme se traduira par la suite d'événements suivants : la voyelle à la fin est entourée par deux consonnes, donc une syllabe CVC est récupérée et la séquence consonne-voyelle restante est CV.CV.CV. La séquence CVCVCV se termine par une voyelle avec une consonne à sa gauche, donc une syllabe de type CV est récupérée. La séquence CVCV restante a une voyelle à son extrémité et une consonne sur la gauche. Encore une fois une syllabe de type CV est récupérée. La séquence CV restante est une syllabe seule de type CV. Après segmentation, pré-traitement et analyse des mots exceptionnels ou des mots à orthographe irrégulière, les règles de suppression d'élision et du [suku:n] sont appliquées en premier. Ces règles génèrent effectivement des sons nuls et peuvent être considérées comme un pré-traitement pour le processus de génération *Graphème-Phonème*, avant l'application de notre jeu ordonné de règles. Les règles de suppression d'élision et [suku:n] sont ensuite suivis par les autres règles non ordonnées : règle [ja maqsu:ra], règle d'insertion d'arrêt glottal, les règles [tanwi:n], la règle du signe de gémination et les règles de ligature. Les règles phonologiques restantes sont classées et appliquées dans l'ordre suivant :

- Les règles Chamsi et Kamari et les règles de ligature sont appliqués en suite ;
- Les voyelles longues sont générées en fonction du contexte des voyelles courtes;
- Les graphèmes restants des voyelles courtes sont remplacés par leurs prononciations respectives;
- Les règles des diphtongues et les règles Graphème-Phone sont appliquées en fin de compte (après syllabation).

Les règles sont structurées donc de la plus particulière à la plus générale, contrairement aux autres langues, comme le Français et l'Anglais, où la création des structures arborescentes est indispensable, vu le nombre considérable de règles [147]. En mettant en œuvre l'algorithme pour l'application des règles, les graphèmes des mots d'entrée sont conservés dans une mémoire tampon. Les mots sont lus séquentiellement à partir de la mémoire tampon et chaque mot est parcouru de gauche à droite en appliquant l'ensemble des règles. Indiquons ici que les mots lus sont d'abord convertis de leur écriture graphémique en AS, en leur écriture graphémique en translittération Buckwalter. Une fenêtre focus (une fenêtre glissante axée sur le graphème en cours d'analyse, y compris son contexte droit « Y » et son contexte gauche « X ») est utilisée pour examiner le mot actuellement en cours de traitement. La taille de la fenêtre est initialement fixée au nombre de graphèmes pour la règle la plus large, *M*. La fenêtre est initialement positionnée au début du mot courant. *M* graphèmes sont lus à partir du mot et les règles ordonnées sont appliquées à la recherche d'une correspondance. Si une correspondance est trouvée, les phonèmes correspondants sont concaténés en une chaîne-phonèmes de sortie et la fenêtre focus est décalée par *M* graphèmes vers la droite. Le balayage se poursuit à partir de cette position. Si aucune correspondance n'est trouvée, la taille de la fenêtre focus est décrétementée par un, la fenêtre est déplacée au début du mot et *M-1* graphèmes sont lus à partir de ce mot et les règles sont analysées de nouveau jusqu'à ce qu'une correspondance d'application d'une ou de plusieurs règles soit trouvée. Il est à noter qu'une correspondance est tenue d'être trouvée puisque des graphèmes simples (uniques) sont associés à des phonèmes et ont toujours leurs prononciations. Lorsque la taille de la fenêtre focus atteint la valeur 1, la fin du mot courant est atteinte et toutes les règles auront été épuisées, et la taille de la fenêtre est remise à sa valeur maximale *M* graphèmes, un autre mot sera lu à partir de la mémoire tampon et le processus se- réitère.



```

// Algorithme Syllabation_Diphthongue_Generation
// Syllabation
Commencez par le nœud le plus à droite de la chaîne graphémique terminée par un NULL
Repeat
Lire le nœud suivant à gauche ;
If (le nœud contient une voyelle)
{
Recherchez à gauche et à droite du nœud;
A partir des syllabes ( de la configuration la plus longue (CVCC) à la configuration la
plus courte CV);
Chercher une correspondance avec l'un des six types de syllabes;
Sauter un nœud à gauche de la voyelle et insérer un nœud avec un type de données
unique comme Étiquette syllabique;
}
Until une position de nœud avant le plus à gauche de la fin de la liste

// Génération de Diphthongues
Commencez par le nœud le plus à gauche ou le plus à droite de la liste phonémique
syllabiquement étiquetée ;
Repeat
Parcours de la liste (de droite à gauche ou de gauche à droite) à la recherche de syllabes
de type CVC ou CV: C;
If (syllabe de type CVC ou CV :C)
{
Vérifier la consonne de fermeture syllabique
If (la consonne de fermeture est /w/ ou /j/)
{
Supprimer de la liste le nœud contenant la consonne de fermeture ;
Insérer dans la liste le nœud contenant la diphthongue correspondante ;
}
}
Until la fin de la liste
Appel à l' algorithme de Conversion Phonème_Allophone

```

Pseudo-code pour la Syllabation et la Génération de Diphthongues

3.3 Implémentation

Cette section présente la méthodologie d'implémentation du système phonétiseur TOP-AS : langage de programmation adopté, le type d'encodage retenu, et la mise en œuvre finale.

3.3.1 Choix du langage de programmation

Une décision critique de conception quant au choix du langage de programmation. En effet, nous étions confrontés ici à deux options concurrentielles : *Java* et *python*.

Points pour java :

- Il possède des packages disponibles à utiliser pour la manipulation d'un corpus arabe. Ainsi, des codes pré-existants peuvent être réutilisés.
- L'utilisation de Java et de certains de ses packages disponibles signifie la possibilité d'intégrer ce projet dans des systèmes « *Open Source* » de synthèse de la parole pour l'AS.

Points contre Java :

- Le code existant d'exploitation des corpus en AS est compliqué et nécessite donc du temps et beaucoup d'efforts pour être compris et réutilisé ensuite dans l'implémentation de ce projet.

Points pour Python :

- plus confortable à manier ;
- plus rapide dans la phase de mise en œuvre et très lisible par l'Homme ;
- Il prend correctement en charge le décodage et l'encodage des caractères arabes ;
- Nous ne savons pas à l'avance les problèmes qui pourraient surgir dans cette étape cruciale d'implémentation de notre système de transcription, et combien de temps faudrait-il pour être résolue, il est donc plus sûr de choisir un moyen rapide d'implémentation et ajouter plus de fonctionnalités, si le temps le permet, plutôt que de faire face à deux problèmes : le problème d'implémentation, en plus de l'apprentissage du langage de programmation à utiliser.

Points contre python :

- lent dans le traitement, vu son statut de langage de programmation de haut niveau. Cependant, ceci ne devrait pas être une contrainte majeure, car l'exécution du programme ne sera que pendant la mise en œuvre à des fins de test. En d'autres termes, le système n'est ni exécutable de façon régulière, ni prévu pour un usage personnel de l'utilisateur.
- si le système est construit avec python, il sera alors difficile de l'intégrer aux systèmes construits autour de Java.

Après cette discussion, il a été décidé d'utiliser python. Et après avoir fait ce choix, une recherche a été effectuée pour déterminer quelle version de python à utiliser. Notre choix s'est porté sur sa version 3.5, car elle prend en charge correctement les caractères arabes.

3.3.2 Type d'encodage/Décodage

Notre système de transcription phonétique pour l'AS (TOP-AS) sera donc implémenté en langage python. La lecture du texte arabe se fait en mode 16 bits (en système Hexadécimal) à cause de la norme Unicode. L'avantage de l'utilisation de cette norme est la lecture directe du texte arabe sans avoir besoin de configurer la machine en langue arabe ; ceci assure la portabilité du système de phonétisation (*Tableau 3.11 a*). Nous avons adopté dans notre système de Conversion Graphème-Phonème à base de règles, le système de translittération Buckwalter pour représenter les graphèmes de l'AS (*Annexe 3*) et une version étendue de «Speech Assessment Methods Phonetic Alphabet» (SAMPA) pour représenter les sons de la langue (*voir l'annexe 2*). La raison pour laquelle nous utilisons la notation SAMPA plutôt que de l'IPA est que l'IPA n'est pas un système convivial pour la codification informatique. Il y a un certain nombre de versions alternatives orientées Unicode de l'IPA, mais SAMPA nous convient le plus quand à la possibilité d'intégrer dans notre phonétiseur des parties d'autres logiciels (la boîte à outils du traitement du langage naturel (NLTK), la boîte à outils du modèle de Markov caché (HTK), ou le moteur de synthèse de la parole MBROLA entre autres) qui font usage de cette notation.

Tableau 3.11 a: Standard Unicode pour les caractères de l'AS

Unicode (Hex)	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
062x		ء	آ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
063x		ذ	ر	ز	س	ش	ص	ض	ظ	ع	غ					
064x		ف	ق	ك	ل	م	ن	ه	و	ى	ي	ـَ	ـِ	ـِ	ـِ	ـِ
065x	ـِ	ـِ	ـِ	ـِ												

En termes d'encodage, il faut prendre en considération la structure graphique particulière de l'AS. En effet, les mêmes caractères arabes prennent des formes graphiques différentes en fonction de leur position dans une chaîne de texte et des caractères environnants. Ces traits sont pris en charge par la table UNICODE étendue (Tableau 3.11 b).

Tableau 3.11 b : Codification UNICODE de l'alphabet de base de l'AS selon toutes les formes contextuelles.

General Unicode	Contextual forms				Name
	Isolated	End	Middle	Beginning	
0623 أ	FE83 أ	FE84 أ			'alif
0628 ب	FE8F ب	FE90 ب	FE92 ب	FE91 ب	bā'
062A ت	FE95 ت	FE96 ت	FE98 ت	FE97 ت	tā'
062B ث	FE99 ث	FE9A ث	FE9C ث	FE9B ث	tā'
062C ج	FE9D ج	FE9E ج	FEA0 ج	FE9F ج	ǧīm
062D ح	FEA1 ح	FEA2 ح	FEA4 ح	FEA3 ح	ḥā'
062E خ	FEA5 خ	FEA6 خ	FEA8 خ	FEA7 خ	ḥā'
062F د	FEA9 د	FEAA د			dāl
0630 ذ	FEAB ذ	FEAC ذ			ḍāl
0631 ر	FEAD ر	FEAE ر			rā'
0632 ز	FEAF ز	FEB0 ز			zayn/zāy

0633	FEB1	FEB2	FEB4	FEB3	sīn
س	س	س	س	س	
0634	FEB5	FEB6	FEB8	FEB7	šīn
ش	ش	ش	ش	ش	
0635	FEB9	FEBA	FEBC	FEBB	ṣād
ص	ص	ص	ص	ص	
0636	FEBD	FEBE	FEC0	FEBF	ḍād
ض	ض	ض	ض	ض	
0637	FEC1	FEC2	FEC4	FEC3	ṭā'
ط	ط	ط	ط	ط	
0638	FEC5	FEC6	FEC8	FEC7	ẓā'
ظ	ظ	ظ	ظ	ظ	
0639	FEC9	FECA	FECC	FECB	'ayn
ع	ع	ع	ع	ع	
063A	FECD	FECE	FED0	FECF	ḡayn
غ	غ	غ	غ	غ	
0641	FED1	FED2	FED4	FED3	fā'
ف	ف	ف	ف	ف	

0642	FED5	FED6	FED8	FED7	qāf
ق	ق	ق	ق	ق	
0643	FED9	FEDA	FEDC	FEDB	kāf
ك	ك	ك	ك	ك	
0644	FEDD	FEDE	FEE0	FEDF	lām
ل	ل	ل	ل	ل	
0645	FEE1	FEE2	FEE4	FEE3	mīm
م	م	م	م	م	
0646	FEE5	FEE6	FEE8	FEE7	nūn
ن	ن	ن	ن	ن	
0647	FEE9	FEEA	FEEC	FEEB	hā'
ه	ه	ه	ه	ه	
0648	FEED	FEEE			wāw
و	و	و			
064A	FEF1	FEF2	FEF4	FEF3	yā'
ي	ي	ي	ي	ي	
0647	FEE9	FEEA	FEEC	FEEB	hā'
ه	ه	ه	ه	ه	
0648	FEED	FEEE			wāw
و	و	و			
064A	FEF1	FEF2	FEF4	FEF3	yā'
ي	ي	ي	ي	ي	

0622 آ	FE81 آ	FE82 ل			'alif maddah
0629 ة	FE93 ة	FE94 ة	—	—	Tā' marbūṭah
0649 ى	FEEF ى	FEF0 ى	—	—	'alif maqṣūrah
U+0671 أ	U+FE87 إ	U+FE88 إ	U+FE85 ؤ	U+FE89 ئ	U+FE8A ئ
U+FE8B أ	U+FE8D إ	U+FE8E إ	U+FE80 أ		

3.3.3 Mise en œuvre du système TOP-AS

Le schéma représentatif du système de phonétisation automatique est illustré dans la *figure 3.4*. La démarche que nous avons adoptée pour la réalisation de ce système se décompose en deux phases de traitements linguistiques : La première phase consiste aux opérations de prétraitement largement discutées dans le chapitre précédent : La segmentation, les ponctuations, les espaces, ..., et harmonisation de façon à ce que le texte prétraité ne comporte aucune ambiguïté pour les traitements ultérieurs. La deuxième phase consiste en la phonétisation du texte prétraité en utilisant deux méthodes différentes. La première méthode est fondée sur l'utilisation des lexiques des exceptions (Mots exceptionnels, abréviations, acronymes, noms propres, ... Voir les *Tableaux 3.12, 3.13, 3.14 et tableau 4.4 du chapitre 4*), en introduisant directement la phonétisation correspondante aux mots sans passer par la base de règles de transcription phonétique, ce qui assure plus de rapidité dans le traitement. La deuxième méthode consiste à traiter le reste du texte en utilisant notre base de règles de transcription phonétique. Nous utiliserons la norme Unicode pour le test des graphèmes de la langue arabe. Les règles établies traitent l'ensemble des réalisations graphémiques de la langue et sont au nombre de 44 graphèmes pour enfin obtenir 37 phonèmes (28 consonnes, 6 voyelles, 3 réalisations vocaliques). La structure des règles élaborées est de la forme suivante : chaque graphème est remplacé par un ou plusieurs phonèmes selon son contexte gauche, son contexte droit, ou les deux à la fois. Nous obtenons ainsi le résultat phonétique du texte sans passer par une table de Conversion Graphème-Phonème (la phonétique est incorporée dans chaque règle de la base de règles).

Nous présentons ci-dessous un exemple d'application d'une règle de transcription, comme la **règle de génération de la voyelle longue** [u:]. Cette règle se présente sous la forme :

- "ـ" (damma) \longrightarrow [u:]/X_Y,
X= {tout graphème représentant une consonne}, Y = {« و »}.

Qui se traduit en langage courant que toute consonne avec une « /d'amma/ » suivie d'un « و » (**waw**), la conversion correspondante en phonème serait la voyelle longue « [u:] ». Et, en terme de codage, ceci indique que la voyelle courte damma '◌ْ' (représentée en Unicode par '64F'), suivie de la lettre waw 'و' (représentée en Unicode par '648'), aura pour transcription phonétique le phonème [u:] selon la notation SAMPA adoptée.

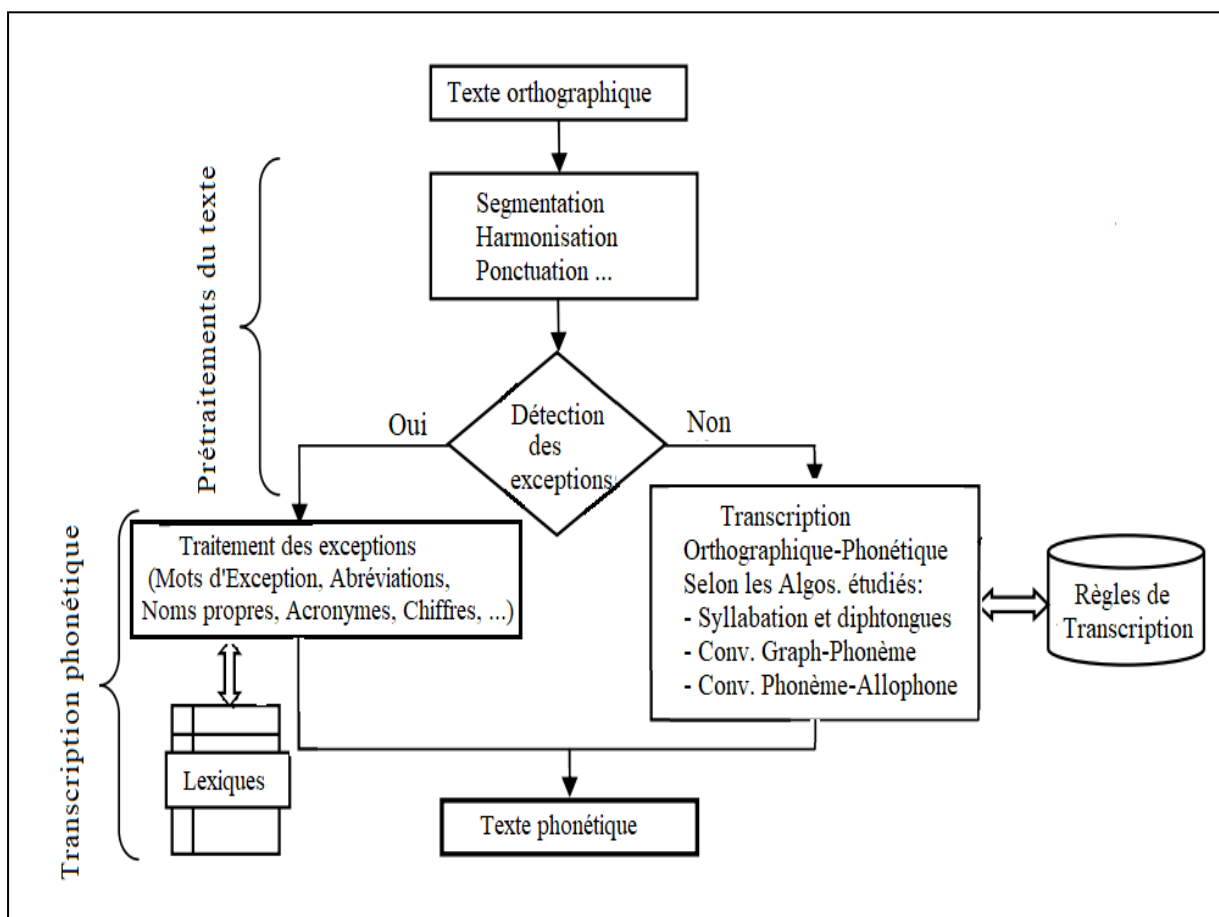


Figure 3.4 : Architecture du système TOP-AS

La première étape consiste à segmenter le texte. Il est primordial dans un premier temps de segmenter le texte donné en phrases, paragraphes, titres etc. En effet, le simple fait d'avoir des ponctuations à chaque fin de phrases n'est pas suffisant, les titres, par exemple, en étant dépourvue. Le résultat de la phonétisation change grandement en fonction de la qualité de la segmentation [148]. Ces entités sont ensuite découpées en unités distinctes, à savoir mots, voire mots composés.

Comme mentionné précédemment, la hiérarchisation des règles est cruciale. Pour obtenir des résultats corrects de notre système TOP-AS, nous avons prévu un séquençement systématique assurant une très haute précision de transcription. Nous avons pris en considération le fait que la sortie d'une règle est l'entrée pour la règle suivante, en s'efforçant de réduire tout écart qui pourrait surgir. La *figure 3.5* illustre l'ordre de ces règles selon des priorités logiques de leurs exécutions : Le processus commence par les règles d'insertion (remplacement d'un graphème par un ou plusieurs phonèmes), puis les règles de conversion et enfin les règles d'élision (suppression).

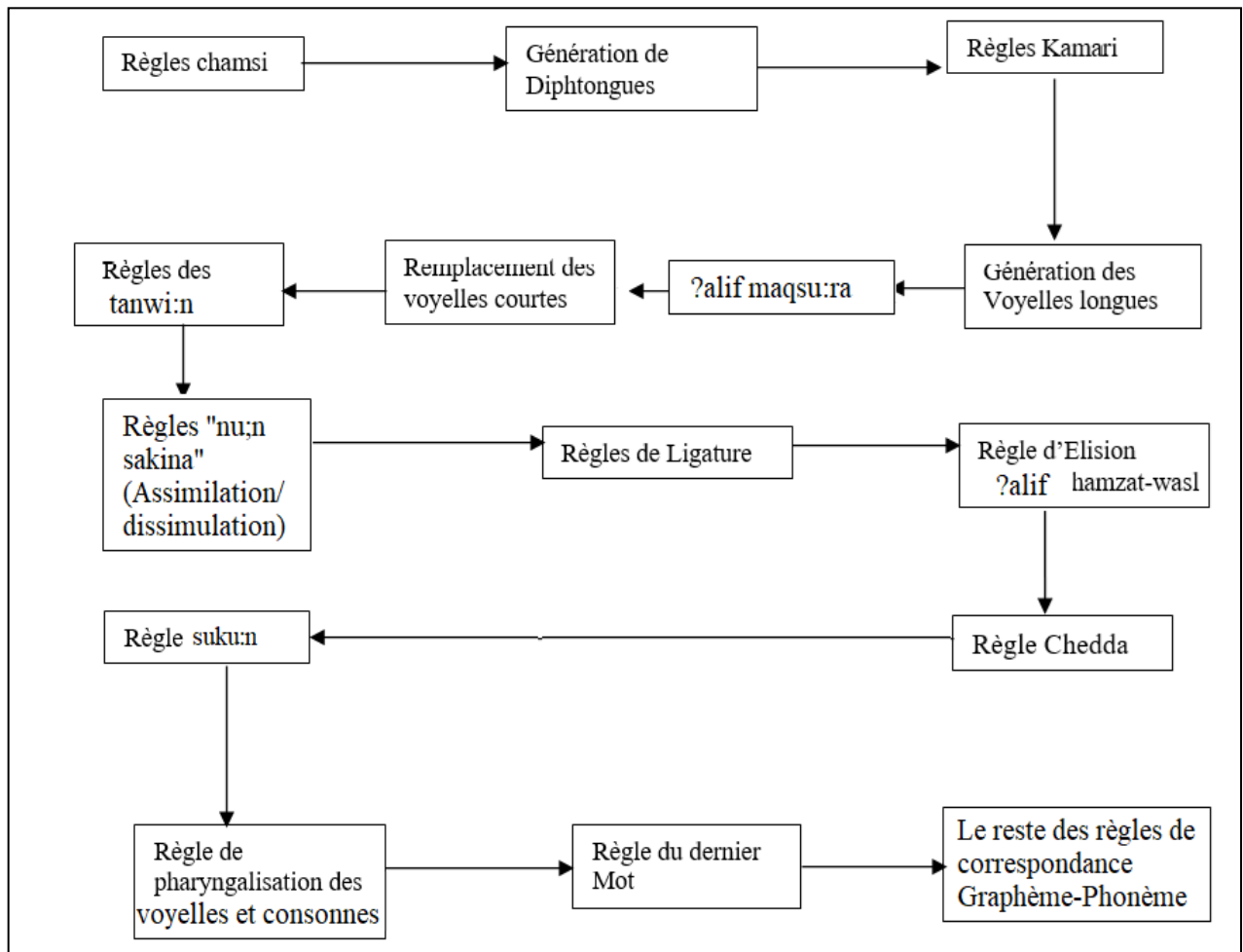


Fig. 3.5 Séquence d'ordonnancement des règles phonétiques et phonologiques du système TOP-AS

3.4 Ressources Numériques

Notre système phonétiseur pour la l'AS, TOP-AS, exploite un certain nombre de ressources comme la Base de Données des abréviations, le lexique des mots d'exception, le lexique des acronymes, symboles et chiffres, compilés et phonétisés par une équipe de doctorants en phonétique de l'Université de Constantine, vérifiés et contrôlés par nos soins, et la liste des mots-outils les plus répandus en AS. Nous donnons dans ce qui suit un aperçu de chacune de ces ressources.

Tableau 3.12 : Extrait du lexique des abréviations courantes en arabe avec leurs transcriptions correspondantes.

N°	Abréviation	Sens	Transcription Phonétique en SAMPA
1	أ.ش.أ	وَكَالَةُ أَنْبَاءِ الشَّرْقِ الأَوْسَطِ المِصْرِيَّةِ	waka:latu ?am ^{ba} :?iSSarqil?awsat'i ?almis'rijatu
2	الناتو	مُنْظَمَةُ جَيْفِ الشَّمَالِ الأَطْلَسِيِّ	munaD'D'amatu xilfiSSama:lil?at'lasi:
3	اليويفا	الإِتِّحَادُ الأَوْرُوبِيُّ لِكِرَةِ القَدَمِ	?al?itixa:dul ?u:rubiju likuratilqadami
4	اه/اه	انْتَهَى	?intaha:
5	اوبك	مُنْظَمَةُ الدُّوَلِ المِصْرِيَّةِ لِلْبِتْرُولِ	munaD'D'amatudduwalilmus'addirati lilbitru:li
6	اونا	وَكَالَةُ الأَنْبَاءِ العُمَانِيَّةِ	waka:latul?am ^{ba} :?il?'uma:nijati
7	إ.ع.م	الإِمَارَاتُ العَرَبِيَّةُ المِتَّحِدَةُ	?al ?ima:ra:tu?'arabiyatulmutaxidatu
8	إب	إِبْلَاح	?ibla:G
9	إلخ	إِلَى آخِرِهِ	?ila: ?a:Xirihi
10	إنج	إِنْجَازٌ	?inʒa:zun
11	أ.أ.	أُسْتَاذٌ	?usta:Dun
12	أ.د	الأُسْتَاذُ الدُّكْتُورُ	?al?usta:Dudduktu:r
13	أ.ش.أ	أَنْبَاءُ الشَّرْقِ الأَوْسَطِ	?am ^{ba} :?uSSarqil?awsat'i
14	دب	دِينَارٌ بَحْرِينِي	di:na:r baxri:ni:
15	دج	دِينَارٌ جَزَائِرِي	di:na:r ʒaza:?'iri:
16	ب.ت.ا	وَكَالَةُ الأَنْبَاءِ الأُرْدُنِيَّةِ	waka:latul ?am ^{ba} :?il?urdunijati
17	بس	بِرْنَامُجٌ سَرْدِي	barna:mazun sardijun
18	بنا	وَكَالَةُ الأَنْبَاءِ البَحْرِينِيَّةِ	waka:latul?am ^{ba} :?ilbaxri:nijah
19	ت	تَرْجَمَةٌ	tarʒamatun
20	ت و ج	التَّجْمَعُ الوَطَنِيُّ الدِّيمُقْرَاطِيُّ	?attazamu?'ulwat'anijuddi:muqra:t'iju
21	تاج	تَجْمَعُ أَمَلِ الجَزَائِرِ	taʒamu?' ?amalilʒaza:?'iri
22	ت ج	التَّلْفَظَةُ الجَزَائِرِيَّةُ	?attalfazatulʒaza:?'irijatu
23	تح	تَحْقِيقٌ	taxqi:q
24	تر	تَرْجَمَةٌ	tarʒamatu
25	ثا	ثَانِيَّةٌ	Ta:nijatun
26	ثنا	حَدَّثَنَا	xaddaTana:
27	ج	الجُزءُ	?alʒuz?u
28	ج ا ل	جِهَازٌ اِكْتِسَابِ اللُّغَةِ	ʒiha:zuktisa:billuGah
29	ج ا	جُمْلَةٌ اِسْمِيَّةٌ	ʒumlatun ?ismijatun
30	ج ت م	جَامِعَةُ التَّكْوِينِ المُنَوَّاصِلِ	ʒa:mi?'atuttakwi:nilmutawa:s'ili
31	ج ت و	جَبْهَةُ التَّحْرِيرِ الوَطَنِيِّ	ʒabhatuttaxri:rilwat'ani:
32	ردمد	الرَّقْمُ الدُّوَلِيُّ المَكْتَبِيُّ للدُّوَرِيَّاتِ	?arraqmudduwalijjilmaktabi: liddawrija:ti
33	ردمك	الرَّقْمُ الدُّوَلِيُّ المَكْتَبِيُّ لِلْكَتُبِ	?arraqmudduwalijjilmaktabi: lilkutubi
34	ج ف	جُمْلَةٌ فِعْلِيَّةٌ	ʒumlatun fi?'lijatun
35	ج.م	جُنْيُهُ مِصْرِي	ʒuni:h mas'ri :
36	ج.م.ع	جُمْهُورِيَّةُ مِصْرِ العَرَبِيَّةِ	ʒumhu:rijatu mis'r ?al?'arabijati
37	جستن	الجَمْعِيَّةُ السُّعُودِيَّةُ لِلْعُلُومِ التَّرْبُويَّةِ وَالنَّفْسِيَّةِ	?alʒam?'ijatussu?'u:dijatulil?'ulu:mittarbawijati wannafsijati
38	ح	حَرْفٌ	xarfun
39	ح ع	حِزْبُ العَمَالِ	xizbul?'uma:li
40	حض	حَاضِرٌ	xa:d'irun
41	خ	خَبْرٌ	Xabarun
42	د	دُكْتُورٌ	duktu :run
43	د	دَقِيقَةٌ	daqi:qatun
44	د.إ	دِرْهَمٌ إِمَارَاتِي	dirhamun ?ima:ra:tijun
45	د.ت	دِينَارٌ تُونِسِي	di:na:run?tu:nusijun

46	د.ت	دُونُ تَارِيخ	du:na ta:ri:X
47	د.ج	دِينَارٌ جَزَائِرِيٌّ	di:na:run zaza:?'irjjun
48	د.ط	دُونُ طَبْعَةٍ	du:na t'ab?'ah
49	د.ع	دِينَارٌ عِرَاقِيٌّ	di:na:run ?'ira:qijjun
50	د.م	دِرْهَمٌ مَغْرِبِيٌّ	dirhamum ^m maGribijjun
51	ر.ض	رَضِيَ اللهُ عَنْهُ	rad'jallahu ?'anhu
52	ر.س	رِيَالٌ سُعُودِيٌّ	rija:lun su?'u:dijjun
53	ز.م	زَمَانٌ	zama:nun
54	س	سُؤَالٌ	su?a:lu
55	س	اسْمٌ	?ismun
56	س.ت	سِجِلٌ تِجَارِيٌّ	si:zilun ^t iza:rijjun
57	س.ا	سَاعَةٌ	sa:?'atun
58	س.ا.نا	الْوَكَالَةُ الْعَرَبِيَّةُ السُّورِيَّةُ لِلْأَنْبَاءِ	?alwaka:latul?'arabijjatussu:rijatu lil?am ^{na} :?i
59	س.ب.أ	وَكَالَةُ الْأَنْبَاءِ الْيَمَنِيَّةِ	waka:latul?am ^{na} :?iljamanijati
60	س.ل	سَنْتِلِتْر	san ^t ilitr
61	س.م	سَنْتِيْمِتْر	san ^t i:mitr
62	س.و.نا	وَكَالَةُ الْأَنْبَاءِ السُّودَانِيَّةِ	waka:latul ?am ^{na} :?issu:da:nijati
63	ش.و.ت.ك	الشَّرْكَةُ الْوَطْنِيَّةُ لِلتَّبْعِ وَالْكِبْرِيَّتِ	?aSSarikatulwat'anijatu littabGi wa:lkibri:t
64	ش.ت	شَرِكَةُ تَضَامَن	Sarikatu tad'a:mun
65	ش.ت.ب	شَرِكَةُ التَّوَصِيَّةِ الْبَسِيطَةِ	Sarikatuttaws'ijatilbasi:t'ati
66	ش.ت.م	شَهَادَةُ التَّعْلِيمِ الْمُتَوَسِّطِ	Saha:datutta?'li:milmutawas's'it'i
67	ش.ذ.م.م	شَرِكَةُ ذَاتِ مَسْئُولِيَّةٍ مَحْدُودَةٍ	Sarikatun ḏa:tu mas'u:lijatin maxdu:datin
68	ش.م	شَرِكَةُ مُسَاهَمَةٍ	Sarikatum ^m musa:himatun
69	ش.م	شَرِكَةُ مَخَاصِنَ	Sarikatu maxa:s'atu
70	ش.و.ن.س.ح	الشَّرْكَةُ الْوَطْنِيَّةُ لِلنَّقْلِ بِالسِّكِّ الْخَدِيدِيَّةِ	?aSSarikatulwat'anijatu linnaqli bissikakilxadi:dijati
71	ص	صَفْحَةٌ	s'afxatun
72	ص.ب/ص.ب.	صُنْدُوقٌ بَرِيدٌ	s'u ⁿ du:qu bari:d
73	ص.ل.ع.م	صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ	s'ala:llahu ?'alajhi wa sallam
74	ط	طَبْعَةٌ	t'ab?'ah
75	ظ	ظَرْفٌ	D'arfun
76	ع	عَدَدٌ	?'adad
77	غ	غَرَامٌ	Gara:m
78	ف	فِعْلٌ	fi?'lun
79	فا	فَاعِلٌ	fa:?'ilun
80	في.فا	الْإِتِّحَادُ الدَّوْلِيُّ لِكُرَةِ الْقَدَمِ	?al?'ittixa:dudduwali: likuratilqadami
81	ق	قَيْدٌ	qajdun
82	ق	قَرْنٌ	qarnu
83	ق.م	قَبْلَ الْمِيلَادِ	qablalmi:la:di
84	ق.هـ	قَبْلَ الْهَجْرَةِ	qablalhi:zrati
85	ق.ق.ج	قِصَّةٌ قَصِيرَةٌ جَدًّا	qis's'atun qas'i:raturun ziddan
86	ق.م.أ	قَاسِمٌ مُشْتَرِكٌ أَكْبَرُ	qa:simum ^m muStarakun akbar
87	ق.نا	وَكَالَةُ الْأَنْبَاءِ الْقَطْرِيَّةِ	waka:latul?am ^{na} :?ilqat'arijati
88	ك.غ	كِيلُوغْرَامٌ	ki:lu:Gra:m
89	ك.م	كِيلُوْمِتْرٌ	ki:lu:mitr
90	ك.و.نا	وَكَالَةُ الْأَنْبَاءِ الْكُوَيْتِيَّةِ	waka:latul?am ^{na} :?ilkuwajtijah
91	ل	لِتْرٌ	litrun
92	ل.س	لِيْرَةٌ سُورِيَّةٌ	li:raturun su:rijaturun
93	ل.م.د	لَيْسَانَسٌ مَاسْتَرٌ دُكْتُورَاهُ	li:sa:ns ma:star duktu:ra:h

94	م.	مُجَلِّدٌ	maʒalad
95	م	مِيَلَادِيَّةٌ	mi:la:dijah
96	م	مِترٌ	Mitr
97	م ف	مُرَكَّبٌ فِعْلِيٌّ	murakkabun fi?'lijjun
98	م.ت.ف	منظمة التحرير الفلسطينية	munaD'amatuttaxri:rilfilist'i:nijah
99	م.ع.س	المملكة العربية السعودية	?almamlakatul?'arabijatussu?'u:dijah
100	ماب	وَكَالَةُ المَغْرِبِ العَرَبِيِّ لِلأَنْبَاءِ	waka:latulmaGribil?'arbi: lil'am ^{ba} :?i
101	متق	متفق عليه	mutufuqun ?'alajhi
102	مح	مِحْوَرٌ	mixwarun
103	مس	مُسْتَقْبَلٌ	mustaqbalun
104	مض	مَاضٍ	ma:d'in
105	مف	مَفْعُولٌ	maf?'u:lun
106	مك	مَكَانٌ	maka:nun
107	ملم	مِلِمِترٌ	mili:mitr
108	منف	مُنَافِدٌ	munafidun
109	ن ق	نِصْفُ قَرْنٍ	nis'fu qarn
110	ناسا	الإدارة الوطنية للملاحة الفضائية والفضاء	?al?ida:ratulwat'anijatu lilmila:xatilfad'a:?ijati wa:lfad'a:?
111	ننا	الوكالة الوطنية للإعلام في لبنان	?alwaka:latulwat'anijatu lil ?i?'la:mi fi: lubna:n
112	هـ	هِجْرِيَّةٌ	hizrijah
113	و.م.أ	الولايات المتحدة الأمريكية	?alwila:ja:tulmutaxidatul'amri:kijatu
114	واج	وَكَالَةُ الأَنْبَاءِ الجَزَائِرِيَّةِ	waka:latul'am ^{ba} :?ilʒaza:?irijati
115	واس	وَكَالَةُ الأَنْبَاءِ السُّعُودِيَّةِ	waka:latul'am ^{ba} :?issu?'u:dijati
116	واع	وَكَالَةُ الأَنْبَاءِ العِرَاقِيَّةِ	waka:latul'am ^{ba} :?i l?'ira:qijati
117	وال	وَكَالَةُ الأَنْبَاءِ اللِّبْيِيَّةِ	waka:latul'am ^{ba} :?illi:bijati
118	وام	وَكَالَةُ أنْبَاءِ الإمارات	waka:latul ?am ^{ba} :?il?ima:ra:ti
119	وفا	وَكَالَةُ فَلَسطِينَ لِلأَنْبَاءِ	waka:latu falast'i:n lili'am ^{ba} :?i
120	وما	الوكالة الموريتانية للأنباء	?alwaka:latulmu:ri:ta:nijatu lil'am ^{ba} :?i
121	ثنا	حَدَّثْنَا	xadaTana:
122	الخ	إلى آخره	?ila: ?a:Xirihi
123	رحه	رَحِمَهُ اللهُ	raximahulla:hu

Tableau 3.13: Extrait du lexique des Mots d'exception en arabe avec leurs transcriptions correspondantes.

Num	Mot	Prononciation	Transcription Phonétique en SAMPA
1	التي	اللتى	?allati:
2	الذي	الذلى	?allaDi:
3	الذين	الذلىن	?allaDi:na
4	الرّحمن	الرّحمان	?arraxma:n
5	السّموات	السّمّوات	?assama:wa:t
6	الله/الله/الله	اللّاه/اللّاه/اللّاه	?alla:h
7	المائة	المىة	?almi?ah
8	المائتان	المىتان	?almi?ata:n
9	إله	إلاه	?ila:h
10	أربعمائة	أربعمىة	?arba?'umi?ah
11	اسحق	إسحاق	?isxa:q
12	أولئك	أولانك	?ula:?ika
13	بِاسْمِ	بِسمِ	bismi
14	تسعمائة	تسعمىة	tis?'umi?ah
15	ثلاثمائة	ثلاثمىة	Tala:Tumi?ah
16	ثمانمائة	ثمانمىة	Tama:numi?ah
17	حم	حامىم	xa:mi:m
18	خمسائة	خمسمىة	Xamsumi?ah
19	داود	داوود	dawu:d
20	ذلك	ذالِك	Da:lika
21	ذلكم	ذالِكُم	Da:likum
22	ذلكما	ذالِكُما	Da:likuma:
23	ذلكن	ذالِكُنَّ	Da:likunna
24	سبعمائة	سبعمىة	sab?'umi?ah
25	ستمائة	ستمىة	sittumi?ah
26	طاوس	طاووس	t'a:wu:s
27	طه	طاها	t'a:ha:
28	فيم	فىما	fi:ma:
29	كذلك	كذالك	kaDa:lika
30	لكن	لاكن	la:kin
31	لكن	لاكن	la:kinna
32	مائة	مىة	mi?ah
33	مائتان	مىتان	mi?ata:n
34	ههنا	هاهنا	ha:huna:
35	هأنا	ها أنا	ha:?ana:
36	هأنتم	ها أنتم	ha:?an'tum
37	هأنذا	ها أنذا	ha:?anaDa:
38	هذا	هاذا	ha:Da:
39	هذان	هاذان	ha:Da:n
40	هذه	هاذيه	ha:Dihi
41	هذين	هاذين	ha:Dajn
42	هكذا	هاكذا	ha:kaDa
43	هكذا	هاكذا	ha:kaDa
44	ههنا	هاهنا	ha:huna:
45	هؤلاء	هاؤلاء	ha:?ula:?
46	يأيتها	يا أئىها	ja:?ajjutuha:
47	يأبها	يا أبها	ja:?ajjuha:
48	يس	ياسىن	ja:si:n

Tableau 3.14 : Extrait du lexique des Acronymes les plus répandus en langues étrangères et qu'on peut rencontrer dans les textes arabes avec leurs sens en arabe et leurs transcriptions correspondantes.

Num	Mot	Sens	Transcription Phonétique en SAMPA
1	adj	صِفَةٌ	s'ifatun
2	a	سَنَةٌ	sanah
3	ADN/DNA	دِي أَنْ آي	di:?ana:j
4	adr.	الْعُنْوَان	?al?'u ⁿ nwa:n
5	AG	جَمْعِيَّة عَامَّة	ʒam?'ijah ?'a:mah
6	AI	الدِّكَاةُ الإصْطِنَاعِي	?aDDaka:?ul?'is't'ina:?i:
7	AM	صَبَاجِيٌّ	s'aba:xijjun
8	ANDRU	الْوَكَاةُ الْوَطْنِيَّةُ لِتَطْوِيرِ الْبَحْثِ الْجَامِعِي	?alwaka:latulwat'anijatu litat'wi:rilibaxTilʒa:mi?'i:
9	ann.	مُلْحَقَةٌ	mulxaqatun
10	ANP	الْجَيْشُ الشَّعْبِي الْوَطْنِي	?aʒʒajSuSSa?'bilwat'ani:
11	ANSEJ	الْوَكَاةُ الْوَطْنِيَّةُ لِدَعْمِ تَشْغِيلِ الشَّبَابِ	?alwaka:latulwat'anijatu lida?'mi taSGi:liSSaba:b
12	APN	الْمَجْلِسُ الشَّعْبِي الْوَطْنِي	?almaʒlisuSSa?'bilwat'ani:
13	AR	عَرَبِيَّة	?'arabijah
14	art.	المَادَّة	?alma:dah
15	av.	شَارِع	Sa:ri?'
16	BAC	بَاكُلُورِيَا	ba:kalu:rja:
17	BBC	هَيْئَةُ الإذَاعَةِ الْبَرِيطَانِيَّة	haj?atul?i:Da:?'atilbari:t'a:nijah
18	BCA	الْبَنْكُ الْمَرْكَزِي الْجَزَائِرِي	?albankulmarkazilʒaza:?iri:
19	BDL	بَنْكُ التَّنْمِيَّةِ الْمَحَلِّيَّة	bankuttanmijatilmaxallijati
20	BEA	الْبَنْكُ الْخَارِجِي الْجَزَائِرِي	?albankulXa:riʒj ?alʒaza:?iri:
21	bibl.	مَكْتَبَةٌ	maktabah
22	BNA	الْبَنْكُ الْوَطْنِي الْجَزَائِرِي	?albankulwat'ani: ?alʒaza:?iri:
23	BP	صُنْدُوقُ بَرِيد	s'u ⁿ ndwqu bari:d
24	BRCS	جَمْعِيَّةُ الصَّلِيْبِ الْأَحْمَرِ الْبَرِيطَانِي	ʒam?'ijatus's'ali:bi:l?axmarilbri:t'a:ni:
25	BTP	الْبِنَاءُ وَالْأشْغَالُ الْعُمُومِيَّة	?albina:?u wal?al?aSGa:lul?'umu:mijah
26	bur.	مَكْتَب	maktab
27	c. c.	صُورَةٌ طَبِيقُ الْأَصْلِ	s'u:ratun t'ibqal?as'l
28	CCP	الْحِسَابُ الْجَارِي الْبَرِيدِي	?alxisa:bulʒa:ri: ?albari:di:
29	CD	قُرْصٌ مَضْغُوط	qurs'um ⁿ mad'Gu:t'
30	CDER	مَرْكَزُ تَطْوِيرِ الطَّاقَاتِ الْمُنْجِدَّة	markazu tat'wi:rilt'a:qa:tilmutaʒaddidati
31	CD-R	مُحْرَكُ نَسْخِ لِقُرْصِ الْمَضْغُوط	muxarik nasX lilqurs'ilmad'Gu:t'
32	CDTA	مَرْكَزُ تَطْوِيرِ التِّكْنُولُوجِيَّاتِ الْمُتَقَدِّمَةِ	markazu tat'wi:rittiknulu:ʒijja:tilmutaʒaddimat
33	chap.	فَصْلٌ	fas'l
34	CHU	مُسْتَشْفَى جَامِعِي	mustaʒfa: ʒa:mi?'i:
35	CIA	وَكَاةُ الإِسْتِخْبَارَاتِ الْأَمْرِيكِيَّة	waka:latul?istiXba:ra:til?amri:kijati

36	C ^{ie}	شَرِكَة	Sarikah
37	CNEP	الصُّنْدُوقُ الوَطَنِي لِلتَّوْفِيرِ وَالِإِخْتِيَابِ	?as's'u ^{ndu} :qulwat'ani: littawfi:ri wal?ixtija:t'
38	CNES	المَجْلِسُ الوَطَنِي الإِقتِصَادِي وَالِإِجْتِمَاعِي	?almaʒlisulwat'ani: ?al?iqtis'a:di: wal ?iztima:?'i:
39	CNR	الصُّنْدُوقُ الوَطَنِي لِلْمَعَاشَاتِ	?as's'undu:qulwat'ani: lilma?'a:Sa:t
40	CO2	ثَانِي أُكْسِيدِ الكَرْبُونِ	Ta:ni: ?uksi:d ?alkarbu:n
41	col.	خَانَةٌ	Xa:natun
42	coll.	مَجْمُوعَة	mazmu:?'atun
43	CP	الرَّقْمُ البَرِيدِي	?arraqmulbari:di:
44	CPA	الْقَرْضُ الشَّعْبِي الجَزَائِرِي	?alqard'uSSa?'bilʒaza:?'iri:
45	CPU	وَحْدَةُ المُعَالَجَةِ المَرْكَزِيَّةِ	wixdatulmu?'a:laʒatilmarkazijah
46	CV	سِيَرَةٌ ذَاتِيَّةٌ	si:ratun Da:tijatun
47	DAS	مُدِيرِيَّةُ الشُّوْنِ الإِجْتِمَاعِيَّةِ	mudi:rijatuSSu?'u:nil?iztima:?'ijah
48	DG	المُدِيرُ العَامِ	?almudi:rul?'a:m
49	DGSN	المُدِيرِيَّةُ العَامَّةُ لِلمَأْمَنِ الوَطَنِي	?almudi:rijatul?'a:mah lil?ammilwat'ani:
50	dict.	قَامُوسٌ	qa:mu:sun
51	dim	الأَخْدُ	?al?axadu
52	Doc	وَثِيْقَة	waTi:qah
53	dom.	الْبَيْتُ	?albajt
54	D ^r , Dr	دُكْتُور	duktu:r
55	DRS	دَائِرَةُ الإِسْتِعْلَامَاتِ وَالْأَمْنِ	da:?'iratul ?isti?'la:ma:t wal?amn
56	éd.	طَبْعَة	t'ab?'ah
57	édit.	النَّاسِرُ	?anna:Sir
58	E-MAIL	البَرِيدُ الإِلِكْتُرُونِي	?albari:dul?ilaktru:ni:
59	EN	إِنْجَلِيزِيَّة	?inzali:zijjah
60	ENA	المُدْرَسَةُ الوَطَنِيَّةُ لِلإِدَارَةِ	?almadrasatulwat'anijjatu lil?ida:rah
61	ENP	المُدْرَسَةُ الوَطَنِيَّةُ المُتَعَدِّدَةِ التَّقْنِيَّاتِ	?almadrasatulwat'anijjatulmuta?'adidatitqnija:t
62	ENSET	المُدْرَسَةُ العُلْمِيَّةُ لِإِسَاتِدَةِ التَّعْلِيمِ التَّقْنِي	?almadrasatul?'ulja: li?sa:tiDatitta?'li:mittiqni:
63	ex.	مِثَال	miTa:l
64	F	مُؤَنَّثٌ	mu?annaTun
65	FAO	مُنْظَمَةُ الأُمَمِ المُتَّجِدَةِ لِلتَّغْذِيَّةِ وَالزَّرَاعَةِ	munaD'D'amatul?umamilmuttaxidah littaGDijati wazzira:?'ah
66	FAQ	أَسْئَلَةٌ مُتَكَرِّرَةٌ مِنْ رُؤَارِ الأَنْتَرْنِيَّتِ	?as?ilatun ^m mutakarriratun min zuwwa:ril?an ^o tarni:t
67	fb	فَيْسْبُوك	fajsbu:k
68	FBI	المَكْتَبُ المُدِيرِي لِلمُتَحَقِّقَاتِ	?almaktabulfdi:ra:li: littaxqa:qa:t
69	fém.	مُؤَنَّثٌ	mu?annaT
70	FFS	جَبْهَةُ القُوَى الإِسْتِرَاكِيَّةِ	zabhatulquwa: ?al?iStira:kijati
71	FLE	الفِرَانْسِيَّةُ كَلْمَةً أِجْنَبِيَّةً	?alfiransijatu kaluGatin ?aʒnabijatin
72	FLN	جَزْبُ جَبْهَةِ التَّخْرِيرِ الوَطَنِي	xizbu zabhatittaxri:r ?alwat'ani:

73	FMI	صندوق النقد الدولي	s'u ⁿ du:qunnaqdidduwali:
74	FR/fr.	فرنسية	firansijjah
75	Fri	الجمعة	?alzumu?'atu
76	FTP	بروتوكول نقل الملفات	bru:tu:ku:l naqlilmilafa:t
77	G7	مجموعة السبعة	mazmu:?'atussab?'ah
78	GATT	الاتفاقية العامة للتعريفات الجمركية والتجارة	?al?ittifa:qijah ?al?'a:mah litti?'ri:fatilzumrukijah watti3a:rah
79	GMT	غرينتش	Gri:nitS
80	GPL	الغاز المميع	?alGa:zulmumajja?'u
81	gr	يونانية	ju:na:nijjah
82	HCI	المجلس الإسلامي الأعلى	?alma3lisul?isla:mi: ?al ?a?'la:
83	HD	القرص الصلب	?alqurs'us's'ulbi
84	hist.	تاريخ	ta:ri:X
85	HLM	سكنات منخفضة الإيجار	sakana:t mu ⁿ Xafid'atul?i:3a:r
86	HT	بدون احتساب الرسوم	bidu:nixtisa:birrusu:mi
87	IAP	المعهد الجزائري للبترول	?alma?'hadul3aza:?'iri: lilbitru:li
88	id.	نفسه	nafsuhu
89	inf.	أسفل	?asfal
90	INI	المعهد الوطني للمعلوماتية	?alma?'hadulwat'ani: lilma?'lu:ma:tijah
91	introd.	مدخل	madXal
92	IP	النظام المستخدم في الشبكات للاتصال مع الإنترنت	?anniD'a:mulmustaXdamu fi: ?aSsabaka:t lil?itis'a:l ma?'al?an ^o tarni:t
93	ISBN	الرقم الدولي الموحد للكتاب	?alraqmuddwali: ?almuwaxxad lilkutub
94	ISP	مزود خدمة الإنترنت	muzawwidu Xidmatil ?an ^o tarni:t
95	ISSN	الرقم الدولي الموحد للدوريات	?arraqmuddwalj: ?almuwaxxad liddawrja:t
96	jeu	الخميس	?alXami:su
97	LADDH	الرابطة الجزائرية للدفاع عن حقوق الإنسان	?arra:bit'atul3aza:?'irijah liddifa:?'i ?'an xuqu:qil?insa:n
98	lat	لاتينية	la:tinijjatun
99	lat.pop	لاتينية عامية	la:tinijjatun ?'a:mijjatun
100	LMD	نظام أل.أم.دي	niD'a:mu ?al ?am di:
101	lun	الاثنين	?al?'Tnajn
102	M	مذكر	muDakkarun
103	M.	السيد	?asajjid
104	mar	الثلاثاء	?aTTula:Ta:?'u
105	max.	أقصى حد	?aqs'a: xaddin
106	MD	ماركة مسجلة	ma:rkatim ⁿ musaz3alah
107	M ^e	الأستاذ	?al ?usta:D
108	mer	الأربعاء	?al?arbi?'a:?'u
109	MESRS	وزارة التعليم العالي	wiza:ratutta?'li:mil?'a:li :
110	min.	أدنى حد	?adna xad

111	M ^{llc}	الْأَيْسَةُ	?al?a:nisatu
112	M ^{mc}	السَّيِّدَةُ	?assajjidatu
113	Mon	الْإِثْنَيْنِ	?al?iTnajn
114	n	إِسْمٌ	?ismun
115	N.	شمال	Sama:l
116	N. B.	مُلاحِظَةُ	mula:xaD'ah
117	Nasa	وَكَالَةُ الْفَضَاءِ الْأَمْرِيكِيَّةِ	Waka:latulfad'a:?il?amri:kijati
118	NLP	مُعَالَجَةُ اللُّغَاتِ الطَّبِيعِيَّةِ	mu'?a:lazatulluGa:tit't'abi:?'ijati
119	N° ou n°	رَقْمٌ	raqm
120	O.	غَرْبٌ	Garb
121	obs.	مُلاحِظَةُ	mula:xaD'ah
122	OGM	الْمَوَادُّ الْمُعَدَّلَةُ وَرَائِيًا	?almawa:dulmu?'addalatu wira:Tijan
123	OLP	مُنْتَظَمَةُ التَّحْرِيرِ الْفَلَسْطِينِيَّةِ	munaD'D'amatuttaxri:rilmfalast'i:nijah
124	OMC	الْمُنْتَظَمَةُ الْعَالَمِيَّةُ لِلتَّجَارَةِ	?almunaD'D'amatul?'a:lamijjatu littiza:rah
125	OMS	الْمُنْتَظَمَةُ الْعَالَمِيَّةُ لِلصِّحَّةِ	?almunaD'D'amatul?'a:lamijjatu lis's'ixati
126	ONDA	الدِّيوانُ الْوِطْنِيُّ لِحُقُوقِ الْمَوْأَفِ	?addiwa:nulwat'anijju lixuqu:qilmu?allif
127	ONU	مُنْتَظَمَةُ الْأُمَمِ الْمُتَّحِدَةِ	munaD'D'amatul?umamilmuttaxidati
128	OPEP	مُنْتَظَمَةُ الدُّوَلِ الْمُصَدِّرَةِ لِلنَّفْطِ	munaD'D'amatuddwalilmus'addiratu linnift'i
129	OPGI	دِيوانُ التَّرْقِيَّةِ وَالتَّسْيِيرِ الْعَقَارِيِّ	di:wa:nuttarqijati wattasji:ril?'aqa:ri:
130	OPU	دِيوانُ المَطْبُوعَاتِ الجامِعيَّةِ	di:wa:nulmat'bu:?'a:tilza:mi?'ijati
131	OTAN	مُنْتَظَمَةُ جُنُودِ الشَّمَالِ الْأَطْلَسِيِّ	munaD'D'amatul xilfiSSama:lil ?at'lasi:
132	OUA	مُنْتَظَمَةُ الْوَحْدَةِ الإفريقيَّةِ	munaD'D'amatulwixdatil ?ifri:qijati
133	OVNI	جِسْمٌ طَائِرٌ مَجْهُولُ الْهُويَّةِ	zismun t'a:?irum ^m mazhu:lulhuwijjati
134	p.	صَفْحَةٌ	s'afxah
135	p. c. /p. cent	بِالْمِائَةِ	bilmi?ati
136	p. ex.	عَلَى سَبِيلِ الْمِثَالِ	?'ala: sabi:lilmiTa:l
137	P.J.	وَيْفَقَ مُرْفَقَةٍ	waTi:qatum ^m murfaqah
138	P.-S.	مُلاحِظَةُ	mula:xaD'ah
139	p.-v.	مَخْضَرٌ	maxd'ar
140	PAF	شُرْطَةُ الْخُدُودِ الْجَزَائِرِيَّةِ	Surt'atulxudu:dilzaza:?irijjah
141	part.	جُزْءٌ	zuz?
142	PC	الْحَاسِبُ الشَّخْصِي	?alxa:su:buSSaXs'i:
143	pdg	الرَّئِيسُ الْمُدِيرُ الْعَامُ	?arra?i:sulmudi:rul?'a:mu
144	pharm.	صَيْدَلِي	s'ajdali:
145	phys.	فِيْزِيَاءٌ	fi:zja:?
146	Pl	جَمْعٌ	zam?'un
147	plur.	جَمْعٌ	zam?'un
148	PM	مَسَائِيٌّ	masa:?ijjun

149	PME	المؤسسات الصغيرة والمتوسطة	?almu?asasa:tus's'aGi:ratu walmutawas's'it'atu
150	PMI	الصناعات الصغيرة والمتوسطة	?as's'ina:?'a:tus's'aGi:ratu wa?almutawas's'it'atu
151	PNB	الناتج القومي الخام	?anna:tizulqawmi: ?alXa:m
152	PPA	جزب الشعب الجزائري	xizbuSSa?'bilzaza:?'iri:
153	P ^r ou Pr	البروفيسور	?albru:fi:su:r
154	prép	أداة جِر	?ada:tu zar,
155	QCM	أسئلة الاختبارات المتعددة	?as?ilatul ?iXtija:ra:tilmuta?'addidah
156	QI	معدل الذكاء	mu?'addaluDDaka:?'i
157	R.S.V.P.	رد من فضلك	rad min fad'lik
158	R.-V./RDV	مؤعد	maw?'id
159	RAM	ذاكرة الدخول العشوائي	Da:kiratudduXu:lil?'aSwa:?'i:
160	RDC	الطابق الأرضي	?at't'a:biqul?ard'i:
161	réf.	مرجع	marzi?'
162	S.	جنوب	zanu:b
163	s.d.	بدون تاريخ	bidu:ni ta:ri:X
164	s.l.	دون عنوان	du:na ?'u ⁿ nwa:n
165	s.l.n.d.	بلا تاريخ ولا مكان	bila: ta:ri:X wala: maka:n
166	S.V.P.	من فضلك	min fad'lik
167	sam	السبت	?assabtu
168	SAMU	مصلحة الإسعافات الطبية الاستعجالية	mas'laxatul?is?'a:fa:tit't'ibijatil?isti?'za:lijati
169	SARL	مؤسسة ذات مسؤوليات محدودة	mu?asasatun Da:tu mas?u:lija:tim ^m maxdu:datin
170	Sat	السبت	?assabtu
171	sc.	علوم	?'ulu:m
172	SDF	مشرّد	muSarrad
173	sing.	مفرد	mufrad
174	SONALGAZ	شركة الكهرباء والغاز	Sarikatulkehruba:?'i walGa:z
175	SONATRACH	سوناطراك	su:na:t'ra:k
176	SPA	شركة ذات أسهم	Sarikatun Da:tu ?ashum
177	S ^{té}	شركة	Sarikah
178	Sun	الأخذ	?al?axadu
179	t.	مجلّد	muZallad
180	T.S.V.P.	استدّر من فضلك	?istadir min ^o fad'lik
181	TD	أعمال موجهة	?a?'ma:lum ^m muwa3zahah
182	tél.	هاتف	ha:tif
183	tél. cell.	هاتف خلوي	ha:tif Xaljawi:
184	Thu	الخميس	?alXami:su
185	Toefl	إختبار الإنجليزية كلغة أجنبية	?iXtiba:ruI?in3li:zijatu kaluGatin ?aznabijah
186	TP	أعمال تطبيقية	?a?'ma:lun ^t tat'bi:qijjah

187	TPS	التِّلْفِزْيُونُ عِبْرَ السَّاتِلِ	?attilifizju:n ?'abrassa:til
188	TTC / t.t.c.	بِاخْتِسابِ كُلِّ الرُّسُومِ	bixtisa:bi kullirrusu:mi
189	TTS	تَرْكِيبُ الْكَلَامِ إِطْلَاقًا مِنَ النَّصِّ	tarki:bulkala:mi ?in°t'ila:qam°minannas'i
190	Tue	الثَّلَاثَاءُ	?aTTula:Ta:?u
191	TVA	رَسْمٌ عَلَى الْقِيَمَةِ الْمُضَافَةِ	rasmun ?'alalqi:matilmud'a:fah
192	TVHD	تَأْفَرُّهُ الدِّقَّةُ الْعَالِيَةُ	talfazatuddiqqatil?'a:lijati
193	UE	الإِتِّحَادُ الْأُورُوبِي	?al?ittixa:dul?u:ru:bi:
194	UFC	جَامِعَةُ التَّكْوِينِ الْمُتَوَاصِلِ	za:mi?'atuttakwi:nilmutawa:s'il
195	UGTA	الإِتِّحَادُ الْعَامُّ لِلْعَمَلِ الْجَزَائِرِيِّ	?al?ittixa:dul?'a:mu lil?'uma:lilzaza:?irijji:na
196	UMA	إِتِّحَادُ الْمَغْرِبِ الْعَرَبِيِّ	?itixa:dulmaGribil?'arabi:
197	Unesco	اليُونِسْكَو	?alju:nisku:
198	Unicef	اليُونَيْسِيف	?alju:ni:si:f
199	URL	مُؤَشِّرٌ مَكَانٌ وَجُودِ الصَّفْحَةِ	mu?aSSiru maka:ni wuzu:dis's'afxati
200	USMA	إِتِّحَادُ الْعَاصِمَةِ	?itixa:dul?'a:s'imati
201	USTHB	جَامِعَةُ هُوَارِي بَوْمَدِينِ لِلْعُلُومِ وَالتَّكْنُولُوجِيَا	za:mi?'atu huwa:ri: bu:madjan lil?'ulu:mi wattivknu:lu:zija:
202	USTO	جَامِعَةُ وَهْرَانَ لِلْعُلُومِ وَالتَّكْنُولُوجِيَا	za:mi?'atu wahra:n lil?'ulu:mi wattivknu:lu:zija:
203	v	فِعْلٌ	fi?'lun
204	v.-p.	نَائِبُ الرَّئِيسِ	na:?iburra?'i:s
205	ven	الجُمُعَةُ	?alzumu?'atu
206	VIH	فَيْرُوسُ فُقْدَانِ الْمُنَاعَةِ	fajru:su fuqda:ni ?almana:?'ati
207	vip	بِطَاقَةُ الْعَضْوِ الْمُمْتَازِ	bit'a:qatul?'ud'wilmumta:zi
208	v°	ظَهْرُ الْوَرَقَةِ	D'ahrulwaraqati
209	vol.	مُجَلَّدٌ	muZallad
210	Wed	الأَرْبَعَاءُ	?al?arbi?'a:?u
211	www	الشَّبَكَةُ الْعَالَمِيَّةُ الْمَوْسَعَةُ	?aSSabakatul?'a:lamiyatulmuwassa?'atu

3.5 Conclusion

Nous avons présenté dans ce chapitre le formalisme de notre phonétiseur TOP-AS, élaboré à partir des règles exhaustives couvrant tous les phénomènes phonétiques et phonologiques de l'AS largement discutées et commentées au chapitre précédent. Nous avons ensuite exposé l'approche adoptée pour la mise en œuvre et l'implémentation de notre système phonétiseur, ainsi que les ressources exploitées dans cette démarche entre Bases de Données et lexiques phonétiques :

- Base de Données des abréviations ;
- lexique des mots d'exception ;
- lexique des acronymes ;
- Base de Données des mots-outils les plus répandus en AS.

Chapitre 4 :
Test et Evaluation
du système TOP-AS

4.1 Introduction

Tous les développements de systèmes, que ce soit dans le milieu industriel ou académique, ont donné lieu à des évaluations. Celles-ci, malgré les nombreuses tentatives de rationalisation, ont été et continuent à être plus ou moins empiriques en particulier dans le domaine du TAP. Quel est l'intérêt de telles évaluations : doivent-elles se situer dans la perspective d'une « compétition » organisée entre différentes applications afin d'en élire la « meilleure » ou sont-elles destinées à analyser en détail les différentes approches existantes, de façon à déterminer les directions de recherche et de développement qui semblent être les plus prometteuses que d'autres ? La question n'est évidemment pas simple dans la mesure où ces deux objectifs ne sont pas totalement contradictoires.

Dans le cas d'une évaluation analytique objective, nous pouvons dresser la liste d'un certain nombre de points qui ne constituent pas des critères à proprement parler, mais correspondent à des informations pertinentes pour la compréhension du fonctionnement d'un système. Pour ce qui concerne les systèmes de TAP, on peut par exemple être intéressé par des aspects de performance bien entendu (pertinence, mais aussi couverture, robustesse, etc.), de qualité d'implantation (réutilisabilité, généralité, adéquation du paradigme d'implantation au problème, algorithmes, heuristiques, etc.), du type de ressources utilisées (lexiques, grammaires, corpus, Bases de Données acoustiques, etc.) ou encore des caractéristiques plus théoriques (théorie linguistique utilisée, architecture du système, etc.). Il ne s'agit bien entendu pas d'une liste exhaustive, mais ces différents aspects sont fondamentaux pour une évaluation en profondeur des systèmes.

L'absence d'un standard méthodologique et le manque de critères objectifs universellement reconnus rendent la tâche d'évaluation de la composante TOP d'un système TTS plus difficile. Dans le présent chapitre, nous survolerons les différentes techniques utilisées pour l'évaluation des phonétiseurs pour l'Anglais, le Français et pour l'AS. Nous présenterons ensuite amplement les ressources de test que nous avons construit pour le besoin d'évaluation de notre système de transcription, puis nous aborderons l'approche adoptée pour l'évaluation de notre système de transcription *TOP-AS*.

4.2 Evaluation de la TOP

Il y a une abondance de littérature sur l'évaluation globale de la qualité de la synthèse de la parole produite par les systèmes TTS [149]. Cependant, il semble y avoir un manque de littérature sur l'évaluation objective de la performance des composants TTS indépendamment les uns des autres, tels que la Conversion Lettre-Son qui est un sous-module essentiel du système de *vocalisation lettre-son*. La Conversion Lettre-Son (ou TOP selon notre propre vocabulaire) trouve son application dans des domaines autres que la synthèse de la parole ; Il est donc essentiel d'évaluer objectivement sa performance sans se référer (ou légèrement) à toute application finale. L'existence de protocoles objectifs de mesures des performances est essentielle au développement des technologies. Deux modes d'évaluations globales des transcritteurs sont en effet couramment utilisés : la comparaison à une prononciation normative unique ; et l'évaluation de systèmes complets. Dans le premier protocole, se pose la question difficile de la détermination non plus d'une, mais d'un ensemble de prononciations de référence : c'est vrai pour le choix des transcriptions possibles d'une phrase, comme pour la définition des prononciations de références pour un nom propre. L'utilisation du second protocole demanderait de savoir mesurer l'impact des erreurs de prononciation sur la qualité globale du système. Toutefois, l'impact de ces erreurs est difficile à séparer de celles des autres composants du système, que ces erreurs soient commises préalablement ou postérieurement à la phase de transcription elle-même.

Les deux mesures que nous avons adoptées pour déterminer les performances du module de conversion TOP sont le taux d'erreur en phonèmes (PER : *Phoneme Error Rate*) et le taux d'erreur en mots (WER : *Word Error Rate*). PER est le pourcentage (%) de phonèmes prédits incorrectement, et WER est le pourcentage (%) de chaînes de phonèmes prédits avec au moins un phonème incorrect. Par conséquent, pour le WER et le PER, les scores inférieurs sont meilleurs, et ceci pour le cas des mots isolés. Pour le cas du texte, en plus de la mesure PER, nous avons utilisé le taux d'erreur en termes de phrases (SER : *Sentence Error Rate*).

La Conversion Lettre-Son peut sembler moins complexe à évaluer que d'autres aspects de la synthèse, la génération de la prosodie par exemple. En effet, on reste dans le domaine du traitement symbolique, des chaînes de caractères. Cependant de nombreuses questions se posent : au préalable, quel alphabet phonétique adopter ? Doit-on évaluer sur un texte (entrée d'un système TTS) ou sur un lexique (qui couvrirait les difficultés des mots de la langue, ou des aspects plus pointus – ex. sigles, noms propres ...) ? Faut-il procéder à une évaluation sur un texte « courant », ou sur un texte construit spécifiquement par un linguiste pour concentrer les difficultés (des irrégularités orthographiques, des phrases phonétiquement riches...) ? Quel est le format d'entrée, vierge ou non de toutes imperfections comme les fautes orthographiques ou typographiques (les problèmes du *hamza*, l'oubli des voyelles longues, ...) ? Quel est le format de sortie, les conventions pour les liaisons ? Quelle métrique poser pour le dépouillement des résultats ?

La présente section vise à apporter des réponses pratiques et réalistes à certaines de ces questions, mais certaines d'autres soulèvent d'autres problèmes théoriques qui sortent du cadre de cette thèse, et qui feront l'objet d'un projet de recherche en collaboration avec des linguistes et des mathématiciens.

4.2.1 Corpus de phonétisation

L'évaluation des phonétiseurs est intimement liée à la tâche visée par cette action. C'est donc le cadre déterminé par la synthèse de la parole qui doit contraindre le choix des domaines pour le corpus de phonétisation. Le module de Conversion Lettre-Son peut être utilisé dans d'autres application du traitement automatique du langage parlé ou écrit (comme la correction orthographique, la reconnaissance de la parole etc.), mais nous nous restreignons ici aux objectifs de la synthèse de la parole.

Même dans le cadre strict, les difficultés de la phonétisation apparaissent liées à des tâches diverses. Nous avons identifié les domaines suivants pour la phonétisation qui influenceront le choix des corpus de test :

- *Lecture de journaux et documents* : C'est depuis des années l'un des domaines privilégiés de la synthèse de la parole. Le but est souvent d'offrir une assistance aux personnes mal ou non-voyants, pour la lecture de documents divers. Les contraintes en sont la nécessité d'une très large couverture (les textes sont à priori illimités et contiennent des exceptions nombreuses, des erreurs, ils évoluent sans cesse). Par contre une certaine tolérance est acceptable pour la prosodie, et l'aspect naturel de la synthèse ;
- *Annuaire inverse* : Ici le domaine est très spécifique : noms propres, adresses, chiffres et numéros doivent être très soigneusement réalisés ;
- *Dialogue homme-machine* : C'est un domaine où la synthèse peut-être générée à partir de concepts plutôt que de textes. La qualité sonore est de première importance. Dans ce domaine, la synthèse de la parole à partir du texte est en concurrence avec la synthèse par concaténation d'unités préenregistrées, lorsque le vocabulaire est limité.
- *Messagerie vocale* : La lecture de courrier électronique, ou d'autres messages pose de problèmes nouveaux. Ce domaine est sans doute appelé à se développer. Il implique la gestion des fautes de frappe ou d'orthographe, et d'un style où l'on peut souvent qualifier « d'oral écrit » : la syntaxe des courriers électroniques semble plus proche de celle de l'oral, et le style est souvent très « relâché » nécessitant un traitement spécialisé.
- *Autres domaines d'application* : Il est certain que le développement du document électronique, de bases de données textuelles et commerciales, la généralisation de la micro-informatique fera ou fait déjà apparaître de nouveaux besoins spécifiques en synthèse de la parole à partir du texte.

- *Lexique de la langue* : En plus de ces tâches, il est souhaitable de confronter les phonétiseurs à un lexique *orthographique-phonétique* général de la langue, par souci de couverture et de cohérence avec les données phonétiques disponibles. L'inventaire rapide précédent montre que plusieurs types de corpus seraient utiles. En effet, il est impossible de définir un corpus universel à toutes les tâches, de plus la finesse de la phonétisation peut être variable d'une tâche à l'autre. Idéalement, le corpus de texte devrait contenir :
 - Des textes journalistiques ;
 - Des lexiques spécialisés : pour les noms propres, les sigles et acronymes (de tels corpus ne sont pas disponibles pour l'AS) ;
 - Des lexiques généraux ;
 - Des textes de messagerie.

4.2.2 Evaluation de la TOP pour l'Anglais

Différentes méthodes ont été utilisées pour évaluer la composante de prononciation à base de règles. Harrison dans [86] s'est basé -pour évaluer la performance de son jeu de règles pour l'Anglais - sur la pondération fréquentielle et a prévu de prononcer correctement jusqu'à 90% des mots dans un échantillon aléatoire de texte en Anglais.

Bernstein et Nessly dans [150] ont utilisé un sous-ensemble de 1000 mots extrait du corpus Brown, et des scores réalisés allant de 65% (pour le sous-ensemble le plus rare) à 86,8% (pour le sous-ensemble le plus commun) de mots correctement prononcés [151]. Le « *Brown corpus* », est l'un des plus grands corpus écrit et oral de l'Anglais parlé de nos jours. C'est le premier corpus électronique construit principalement à des fins de recherches scientifiques, disponible à l'adresse : <<http://khnt.aksis.uib.no/icame/manuals/brown/>>.

Hunnicut dans [88] a démontré l'évaluation préliminaire de son ensemble de règles en utilisant des sous-ensembles de 200 mots extraits également du corpus Brown et ses scores allant de 66% (pour le sous-ensemble le plus rare) à 100% (pour le sous-ensemble le plus commun) de mots correctement prononcés. Utilisant ces petits morceaux de données de test, Hunnicut a exploité la pondération de fréquence pour estimer la performance de son ensemble de règles sur les mots absents du dictionnaire avec un score de 71% de prononciation correcte. Divay et Vitale ont testé leur nouvel ensemble de règles pour l'Anglais sur un sous-ensemble de 19 837 mots extraits du corpus Brown avec un score de 64,37% mots correctes [90].

Compte tenu de toutes ces méthodes d'évaluation non standard, comment la performance d'un composant de prononciation à base de règles d'un système TTS pour l'Anglais doit être évaluée ? Dans [96], les auteurs ont suggéré une procédure standard à suivre pour l'évaluation de la composante de phonétisation avec des critères suivants :

- les méthodes de phonétisation automatique devraient être testées sur les mêmes grands dictionnaires dans leur intégralité car cela démontre clairement la performance asymptotique de la transcription à base de règles sur un large spectre de données de test ;
- l'utilisation d'une métrique standard commune, telle le scoring (notation) en termes de mots corrects, car celle-ci est plus rigoureuse et sensible que la métrique en termes de phonèmes correctes, puisque les mots sont soit correctes ou non ; et
- l'utilisation d'une liste commune de symboles ou d'un ensemble standard de phonèmes de sortie pour une langue donnée.

Plusieurs approches pour la conversion G2P pour l'Anglais exploitent la base de données *CMUDict* de test contenant des mots avec leurs prononciations phonétiques. *CMUDict* est le « *Carnegie Mellon University Pronouncing Dictionary* », le dictionnaire phonétique de l'Université Carnegie Mellon aux Etats Unis, est un dictionnaire de prononciation lisible par machine open source pour l'Anglais contenant plus de 134 000 mots et leurs prononciations. *CMUDict* est activement maintenu et développé. Ses entrées sont particulièrement utiles pour la reconnaissance et la synthèse de la parole, car il a des correspondances entre les mots et leurs

prononciations⁴ [152]. Le *Tableau 4.1* compare les résultats de la TOP à quelques publications récentes pour l'Anglais. Toutes ces publications utilisent l'ensemble de données CMUDict.

Tableau 4.1 : Comparaison des résultats G2P pour l'Anglais entre plusieurs sources d'étude

Système	PER	WER
Galescu [153]	7.0	28.5
Novak [152]	5.9	24.8
Kheang [154]	5.6	29.4
We [155]	5.5	23.4
Rao [156]	9.1	21.3

4.2.3 Evaluation de la TOP pour le Français

Pour le Français, plusieurs efforts ont été entrepris pour l'évaluation du module de transcription orthographique-phonétique dont nous pouvons citer par exemple les travaux de Brigitte B. et son équipe du laboratoire « Parole et Langage » de l'université de Marseille [157]. Ils ont d'abord construit un corpus, nommé « MARC-Fr - Manual Alignments Reference Corpus for French », entièrement phonétisé et aligné manuellement par un expert phonéticien et déposé sous licence GPL sur la forge SLDR⁵. Les évaluations sont effectuées avec l'outil SCTK de NIST⁶. Il calcule le taux d'erreurs de la phonétisation (Err) qui somme les erreurs de substitution (Sub), de suppression (Del) et d'insertion (Ins). Ils ont obtenu une erreur totale de phonétisation de **13,2%**. C. d'Alessandro et son équipe du laboratoire LIMSI à Orsay, pour l'évaluation globale des systèmes de synthèse de la parole en Français, ainsi que l'évaluation locale de ses différents modules, ont procédé à la construction des corpus et des méthodologies normalisés et concédés à la communauté scientifique et industrielle afin de servir de référence pour l'évaluation des systèmes de synthèse en Français [158]. Pour parvenir non pas à une unique « note » d'évaluation mais plutôt à un diagnostic permettant de mettre à jour les faiblesses et atouts des divers systèmes, il a été décidé de séparer l'évaluation en quatre parties, dont les trois premières correspondent à l'organisation générale des systèmes TTS en modules :

- évaluation de la Conversion Lettre-Son ;
- évaluation de la qualité du codage ;
- évaluation de la prosodie ;
- évaluation des systèmes dans leurs globalités.

Néanmoins, ce qui nous intéresse dans cette campagne d'évaluation est la première tâche, l'évaluation de la TOP. Nous nous intéresserons à l'évaluation des autres aspects de la synthèse, que sont la génération prosodique et la qualité de codage (modification/concaténation acoustique) dans nos futurs travaux de recherche.

⁴ - Weide, R.: The CMU pronunciation dictionary, release 0.7a (2014).

<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

⁵ - Speech Language Data Repository, <http://www.sldr.fr>

⁶ - " The NIST Speech Recognition Scoring Toolkit (SCTK) Version 2.4.0 ", [Page Web],

<http://www.nist.gov/speech/tools>, The NIST Speech Group. National Institute of Standards and Technology (NIST), USA. [Accédé Mars 2018].

4.2.4 Evaluation de la TOP en Arabe Standard

En ce qui concerne la TOP en AS, par rapport à d'autres langues, des recherches limitées ont été menées. Nous pouvons citer des travaux connexes dans ce domaine tels que ceux de Selim et Anbar qui ont développé un système de transcription phonétique basé sur des règles pour le texte arabe ; avec un corpus de test puisé des journaux de 291 mots non-voyellés ; leur système a montré un rapport de précision modéré bien qu'ils aient utilisé un nombre limité de mots testés [159]. El-Imam dans [160] a proposé un système TOP pour le texte arabe en étudiant les propriétés de la phonologie et les règles phonétiques de la langue arabe. Cependant, ces règles n'étaient pas clairement hiérarchisées, de sorte qu'elles pourraient se contredire pendant le déroulement du système et produire des sorties inappropriées, car certaines de ces règles doivent être visitées avant d'autres et la sortie d'une règle précédente sera une entrée pour la règle suivante. M.E. Ahmed dans [161] a utilisé environ 150 allophones et combinaisons voyelles / consonnes. Il a appliqué un ensemble de règles de conversion lettre-son pour simplifier la production de la voix par ordinateur. Ses résultats ont montré que les règles étaient la partie principale de son système de synthèse de la parole en AS. Al-Ghamdi et al. dans [162] ont utilisé les règles de phonologie arabe pour convertir le texte en symboles sonores en énumérant un certain nombre de règles phonétiques et phonémiques avec quelques mots exceptionnels, mais ils n'ont pas mis en œuvre leur système de phonétisation pour tester les performances de leur procédure de TOP à base de règles.

D'un autre côté, de nombreux chercheurs se sont rendu compte que la modélisation de la variation de prononciation peut améliorer les performances du processus TOP. Pour ce faire, ils ont proposé différentes approches, dont l'approche classique consistait à générer un dictionnaire arabe multi-prononciation. Par exemple, F. Biadisy et al. dans [163] ont généré un dictionnaire multi-prononciation utilisant des règles de prononciation combiné au système *MADA* (Morphological Analysis and Disambiguation for Dialectal Arabic), comme outil de désambiguïsation morphologique, pour déterminer la prononciation la plus probable d'un mot donné dans son contexte. La méthode proposée a rapporté une amélioration significative de 4,1% de la précision par rapport au système de référence [164].

Ces règles étaient principalement de transcription directe avec quelques règles pour l'assimilation de «lam» ([*l*]) combiné aux lettres solaires, la conversion de [*n*] en [*m*] lorsqu'il est suivi par la plosive labiale [*b*], et emphatiques avec les voyelles pharyngalisées. L'efficacité de l'utilisation du dictionnaire généré a été testée à l'aide d'un système TTS arabe à grand vocabulaire et a obtenu une précision comparable à celle d'un système TTS Anglais à base de même approche et de même taille de corpus de test. Ce travail a ensuite été mis en œuvre dans de nombreuses autres publications comme dans M. Al-Ghamdi et al. [165]. Ramsay et al. dans [166] ont développé un modèle complet basé sur approche emprunté des techniques de l'intelligence artificielle autour d'une base de connaissances (knowledge database), pour générer automatiquement une transcription phonétique à partir d'un texte donné en arabe standard. Ce modèle est basé sur un ensemble de règles de prononciation dépendant de la langue qui fonctionne sur la conversion du texte arabe entièrement diacritisé en sons réels, avec un lexique pour les mots exceptionnels. Cherifi et Guerti dans [110] ont développé une méthode basée sur des transducteurs à états finis pondérés. Le score était de 3.5% en termes de PER et de 12.5% en termes de WER, avec un corpus d'entraînement et de test de 30 000 mots phonétisés. Les mêmes auteurs, dans [78] ont exploité une approche probabiliste avancée basée sur les données, à savoir les champs aléatoires conditionnels (CRF). Même si les résultats ne dépassent pas les meilleurs scores des systèmes de référence existants mais ouvrent la voie vers l'avenir en matière d'innovations dans les techniques « Data-Driven ». En effet, dans tous leurs tests, le score était compris entre 11 et 15% de taux d'erreur sur la transcription des phonèmes (PER). Ces mêmes auteurs, ont proposé dans [77] une approche pour la Conversion Graphème-Phonème en AS, basée sur une autre technique « Data-Driven » : le modèle multi-gramme conjoint (JMM). Dans ce genre d'approches, nous n'avons pas besoin d'expliquer toutes les anomalies de la correspondance Graphème-Phonème, mais toutes les connaissances Orthographico-Phonétiques seront incluses implicitement dans l'étape d'entraînement du modèle. Ils ont obtenus un score d'un peu plus de 10 % en PER pour la transcription de mots isolés, et un score d'un peu plus de 11% sur un corpus de parole continue.

4.3 Evaluation de notre phonétiseur TOP-AS

La langue AS n'est certainement pas aussi complexe que le Français ou l'Anglais et n'a pas les mêmes irrégularités orthographiques que ces deux langues, celles-ci nécessitent de vastes ensembles de règles et de gros lexiques spéciaux. En raison de ces régularités orthographiques, nous avons aménagé et combiné un ensemble de règles suffisamment génériques pour couvrir tous les phénomènes orthographiques, phonétiques et phonologiques de cette langue. Les exceptions (qui relèvent principalement de la catégorie des violations de la génération de la voyelle longue /a:/) aux règles régulières d'orthographe pourraient alors être traitées par un dictionnaire de mots exceptionnels qui doit être analysé avant d'exécuter le jeu de règles génériques. Les seuls points dont nous devons accorder une attention particulière est que les données d'essai doivent être suffisamment complètes et arbitraires pour couvrir toutes les règles, tous les cas exceptionnels et tous les phonèmes (ou phones) à générer autant de fois que possible. On pourrait donc limiter le corpus d'évaluation pour la conversion « *lettre-son* » pour l'AS à un assortiment de mots et de phrases assez riches en contenus phonémiques et phonétiques, qui permettra de tester les principales difficultés dans la transcription *lettre-son*. Ces difficultés ont déjà été largement discutés dans cette thèse (Voir les *chapitres 2 et 3*) et elles comprennent : les structures phonétiques complexes (les règles Graphème-Phonème et Phonème-Phone), les acronymes, abréviations et symboles, les chiffres, les noms propres, la segmentation, etc. Essentiellement, le corpus devrait être riche en contenu phonétique et devrait inclure toutes les variantes de prononciation de la parole arabe.

Une méthode largement répandue pour évaluer la performance de la Conversion Graphème-Phonème serait de comparer sa chaîne de sortie phonémique avec une transcription de référence, puis comptabiliser les écarts comme étant des erreurs de transcription. Une telle approche –dite *évaluation objective*– se concentrerait sur la transcription des mots isolés tirés de dictionnaires phonétiques, puis de prendre des textes de référence en plusieurs domaines, pour tester la performance de transcription des phénomènes phonétiques et phonologiques qui apparaissent entre les bords des mots dans la parole continue.

4.3.1 Corpus d'évaluation

Notre système de transcription lettre-son pour la l'AS que nous avons développé dans cette thèse a été testé sur :

- Un dictionnaire phonétique *MCAW-Dict (Most Commonly-used Arabic Words)* de 35 000 mots les plus fréquemment utilisés en AS, qui ont été compilés par une équipe de linguistes et de phonéticiens du laboratoire du traitement automatique de la langue arabe (LTALA) de l'Université de Tlemcen [167] ;
- Une liste de 25.000 noms propres arabes les plus répons compilés et phonétisés par un groupe d'étudiants en Master II de la filière «Phonétique» du département de la langue arabe de l'université de Tlemcen ; et
- Une compilation de textes courts et diversifiés recueillis de différentes sources et couvrants plusieurs domaines (didactique, technique, académique, littéraire, scolaire et religieux).

Les méthodes à base de règles ne sont pas analogues aux approches d'entraînement « *Data-Driven* » où les données doivent être divisées en deux catégories : Données d'entraînement et données d'essais, et qui sont indépendantes les unes des autres. Une règle est générale, et quand elle est applicable, elle s'applique indépendamment de la classe du mot. Le *Tableau 4.2* montre les occurrences des règles phonologiques et phonétiques appliquées aux mots de test. Les statistiques présentées dans ce tableau montrent qu'un nombre suffisant de mots de test du dictionnaire phonétique *MCAW-Dict* couvre chaque règle correspondant au système orthographique de la langue arabe. Cette liste contient également de nombreux mots arabes irréguliers orthographiquement (mots spéciaux ou mots à orthographe irrégulier). Si nous définissons les « *mots rares* » comme étant les mots en dehors du vocabulaire fréquemment utilisé (en dehors du dictionnaire *MCAW-Dict* pour notre système). En ce qui concerne la relation entre la forme orthographique des mots de l'AS et leurs contenus sonores, les mots rares et les plus fréquemment

utilisés partagent la propriété commune qu'ils obéissent aux règles lettre-son, tant qu'ils ne sont pas des mots ou des noms exceptionnels comme discuté dans la thèse. Ceci est vrai pour l'AS : tout mot arabe est soumis aux mêmes règles d'orthographe et de prononciation génériques tant qu'il n'est pas une exception. Il y a, bien sûr, des mots exceptionnels qui sont rarement utilisés et d'autres qui sont utilisés plus fréquemment, ce qui implique que le concepteur du dictionnaire des exceptions doit veiller à ce que sa liste soit aussi complète que possible, ce que nous avons tenté de faire pour ce système. La liste des mots exceptionnels que nous avons adoptés proviennent des deux lexiques les plus complets pour l'AS [83, 132].

Tableau 4.2 Occurrences des règles phonologiques et phonétiques appliquées aux unités du lexique de test (Dictionnaire Phonétique MCAW-Dict)

Type de règle		Nombre approximatif des occurrences
Règles phonémiques	Suppression du suku:n	7.200
	Élision du Alef ^a	1.950
	Ya maqsu:ra	400
	Insertion de coup glottal	6.000
	[tanwi:n] ^b	12.000
	Gémination	9.000
	Ligatures	4.500
	Chamsi et Kamari	12.000
	Vocalisation (Maad)	14.000
	Diphthongues	2.500
Voyelles brèves (fatha, damma et kasra)	5.000	
Règles phonétiques	Pharyngalisation des voyelles	2.500
	Nasalisation des voyelles	4.350
	Emphase (L'accent) du / ℓ / et / r /	300
	Assimilation du son due à l'emphase	340
	Recouvrement et adaptation	200

^a Obtenu en convertissant environ 500 verbes de leurs formes passé-singuliers aux différentes formes plurielles et comprenant l'élision de *hamzat wasl*.

^b La plupart des mots dans le dictionnaire phonétique *MCAW-Dict* se terminent avec [tanwi:n] Damah. Les terminaisons sont modifiées pour refléter les [tanwi:n] /fatxa/ ou /kasra/.

Tableau 4.3 : Extrait du dictionnaire phonétique de test MCAW-Dict

Num	Mot	Transcription	Num	Mot	Transcription	Num	Mot	Transcription
1	إِبْتَأَسَ	?ibta?asa	31	إِبْيَضَ	?ibjad'd'a	61	إِئْفَاقِيَّة	?ittifa:qi:jah
2	إِبْتَدَأَ	?ibtada?a	32	إِتَّبَعَ	?ittaba?'a	62	إِئْفَاءٌ	?ittifa:?'un
3	إِبْتَدَرَ	?ibtadara	33	إِتَّجَرَ	?ittazara	63	إِتِّكَاءٌ	?ittika:?'un
4	إِبْتَدَعَ	?ibtada?'a	34	إِتَّجَهَ	?ittazaha	64	إِتِّكَالٌ	?ittika:lun
5	إِبْتَدَلَ	?ibtaDala	35	إِتَّخَدَ	?ittaxada	65	إِتِّهَامٌ	?ittiha:mun
6	إِبْتَزَّ	?ibtazza	36	إِتَّخَذَ	?ittaXaDa	66	إِتْنَيْنٌ	?iTnajni
7	إِبْتَسَمَ	?ibtasama	37	إِتَّسَخَ	?ittasaXa	67	إِجْتَبَى	?iztaba:
8	إِبْتَعَثَ	?ibta?'aTa	38	إِتَّسَعَ	?ittasa?'a	68	إِجْتَثَّ	?iztaTTa
9	إِبْتَعَدَ	?ibta?'ada	39	إِتَّسَقَ	?ittasaqa	69	إِجْتَدَبَ	?iztaDaba
10	إِبْتَعَى	?ibtaGa:	40	إِتَّسَمَ	?ittasama	70	إِجْتَرَأَ	?iztara?a
11	إِبْتَكَرَ	?ibtakara	41	إِتَّصَفَ	?ittas'afa	71	إِجْتَرَفَ	?iztarafa
12	إِبْتَلَعَ	?ibtala?'a	42	إِتَّصَلَ	?ittas'ala	72	إِجْتَرَمَ	?iztarama
13	إِبْتَلَى	?ibtala:	43	إِتَّصَحَّ	?ittad'axa	73	إِجْتَرَّ	?iztarra
14	إِبْتَهَجَ	?ibtaha?za	44	إِتَّعَطَّ	?itta?'aD'a	74	إِجْتَرَأَ	?iztaza?a
15	إِبْتَهَلَ	?ibtahala	45	إِتَّفَقَ	?ittafaqa	75	إِجْتَلَبَ	?iztalaba
16	إِبْتَدَاءٌ	?ibtida:?'un	46	إِتَّفَدَ	?ittaqada	76	إِجْتَلَى	?iztala:
17	إِبْتِدَاعٌ	?ibtida:?'un	47	إِتَّقَى	?ittaq:	77	إِجْتَمَعَ	?iztama?'a
18	إِبْتِدَاعِيٌّ	?ibtida:?'i:jun	48	إِتَّكَأَ	?ittaka?a	78	إِجْتَنَبَ	?iztanaba
19	إِبْتِدَائِيٌّ	?ibtida:?'i:jun	49	إِتَّكَلَ	?ittakala	79	إِجْتَنَى	?iztana:
20	إِبْتِدَالٌ	?ibtiDa:lun	50	إِتَّهَمَ	?ittahama	80	إِجْتَهَدَ	?iztahada
21	إِبْتِزَارٌ	?ibtiza:zun	51	إِتِّبَاعٌ	?ittiba:?'un	81	إِجْتَوَى	?iztawa:
22	إِبْتِزَارِيٌّ	?ibtiza:zi:	52	إِتِّجَارٌ	?ittiga:run	82	إِجْتِثَاتٌ	?iztiTa:Tun
23	إِبْتِسَامَةٌ	?ibtisa:matun	53	إِتِّجَاهٌ	?ittiza:hun	83	إِجْتِرَاحٌ	?iztira:xun
24	إِبْتِعَادٌ	?ibti?'a:dun	54	إِتِّخَادٌ	?ittixa:dun	84	إِجْتِرَارٌ	?iztira:run
25	إِبْتِغَاءٌ	?ibtiGa:?'un	55	إِتِّخَادِيٌّ	?ittixa:di:jjun	85	إِجْتِزَاءٌ	?iztiza:?'un
26	إِبْتِكَارٌ	?ibtika:run	56	إِتِّسَاعٌ	?ittisa:?'un	86	إِجْتِلاءٌ	?iztila:?'un
27	إِبْتِلاءٌ	?ibtila:?'un	57	إِتِّسَاقٌ	?ittisa:qun	87	إِجْتِمَاعٌ	?iztima:?'un
28	إِبْتِهَاجٌ	ibtiha:zun	58	إِتِّصَالٌ	?ittis'a:lun	88	إِجْتِمَاعِيٌّ	?iztima:?'i:j
29	إِبْتِهَالٌ	?ibtiha:lun	59	إِتِّضَاحٌ	?ittid'a:xun	89	إِجْتِثَابٌ	?iztina:bun
30	إِبْنٌ	?ibn	60	إِتِّفَاقٌ	?ittifa:qun	90	إِجْتِهَاذٌ	?iztiha:dun

Num	Mot	Transcription	Num	Mot	Transcription	Num	Mot	Transcription
91	إِخْتِيَاخٌ	?i3ti:a:xun	121	إِحْتَقَنَ	?ixtaqana	151	بَرْجَل	bar3al
92	إِخْتِيَازٌ	?i3ti:ja:zun	122	إِحْتَكَّرَ	?ixtakara	152	بَرْد	Bard
93	إِجْتَاخَ	?i3ta:xa	123	إِحْتَكَمَ	?ixtakama	153	بَرْدَعَة	barda?'ah
94	إِجْتَازَ	?i3ta:za	124	إِحْتَكَّ	?ixtakka	154	بَرْزَخ	barzaX
95	إِحْتَبَسَ	?ixtabasa	125	إِحْتَلَبَ	?ixtalaba	155	بَرْطَمَان	bart'ama:n
96	إِحْتَبَى	?ixtaba:	126	إِحْتَلَمَ	?ixtalama	156	بَرْعَم	bar?'am
97	إِحْتَجَبَ	?ixta3aba	127	إِحْتَلَّ	?ixtalla	157	بَرْعَر	barGar
98	إِحْتَجَزَ	?ixta3aza	128	إِحْتَمَلَ	?ixtamaal	158	بَرْغُوث	barGu:T
99	إِحْتَجَّ	?ixta33a	129	إِحْتَمَى	?ixtama:	159	بَرْغُوثِي	barGu:Tijj
100	إِحْتَدَمَ	?ixtadama	130	إِحْتَوَى	?ixtawa:	160	بَرْق	Barq
101	إِحْتَدَّ	?ixtadda	131	إِحْتِيَّاسٌ	?ixtiba:sun	161	بَرْقَة	Barqah
102	إِحْتَرَبَ	?ixtaraba	132	إِحْتِجَابٌ	?ixti3a:bun	162	بَرْقَع	barqa?'
103	إِحْتَرَزَ	?ixtaraza	133	إِحْتِجَاخٌ	?ixti3a:3un	163	بَرْقُوق	barqu:q
104	إِحْتَرَسَ	?ixtarasa	134	إِحْتِجَايِي	?ixti3a:3ijjun	164	بَرْقِيَّة	barqijjah
105	إِحْتَرَفَ	?ixtarafa	135	إِحْتِجَازٌ	?ixti3a:zun	165	بَرْلَمَان	barlama:n
106	إِحْتَرَقَ	?ixtaraqa	136	إِحْتِدَامٌ	?ixtida:mun	166	بَرْمَجَة	barma3ah
107	إِحْتَرَمَ	?ixtarama	137	إِحْتِدَاءٌ	?ixtiDa:?un	167	بَرْمَجِي	barma3ijj
108	إِحْتَرَمَ	?ixtazama	138	إِحْتِرَاؤٌ	?ixtira:zun	168	بَرْمُودَة	barmu:dah
109	إِحْتَسَبَ	?ixtasaba	139	إِحْتِرَازِي	?ixtira:zijjun	169	بَرْمِيل	barmi:l
110	إِحْتَسَى	?ixtasa:	140	إِحْتِرَاسٌ	?ixtira:sun	170	بَرْمَائِي	barma:?ijj
111	إِحْتَسَدَ	?ixtaSada	141	إِحْتِرَافٌ	?ixtira:fun	171	بَرْتُق	Barnaq
112	إِحْتَسَمَ	?ixtaSama	142	إِحْتِرَافِي	?ixtira:fijjun	172	بَرْيَّة	barnijjah
113	إِحْتَضَرَ	?ixtad'ara	143	بَرِيء	bari:?un	173	بَرْنَامَج	barna:ma3
114	إِحْتَضَنَ	?ixtad'ana	144	بَرِيد	bari:dun	174	بَرْهَن	Barhan
115	إِحْتَطَبَ	?ixtat'aba	145	بَرِيدِي	bari:dijjun	175	بَرْهَنَة	barhanah
116	إِحْتَفَرَ	?ixtafara	146	بَرِيْزَة	bari:zatun	176	بَرْوَاز	barwa:z
117	إِحْتَفَظَ	?ixtafaD'a	147	بَرِيْق	bari:qun	177	بَرَّ	barr
118	إِحْتَفَلَ	?ixtafala	148	بَرْبَر	barbar	178	بَرَّد	barrad
119	إِحْتَفَى	?ixtafa:	149	بَرْبَرِي	barbarijjun	179	بَرَّر	barrar
120	إِحْتَقَرَّ	?ixtaqara	150	بَرْبَرِيَّة	barbarijjatun	180	بَرِّي	barrijj

Dans le *Tableau 4.4* suivant, nous présentons un aperçu de la liste des noms propres les plus répandus en AS avec leurs transcriptions en SAMPA. La liste complète contient plus de 25 000 noms. Sur le lien :

https://docs.google.com/spreadsheets/d/1e9xDOj_bD94Gxxk-sAOdD1kS0s2dunNTLISB10FZhTo/edit?usp=sharing

Nous avons hébergé une liste étendue de plus de 500 noms avec leurs transcriptions SAMPA.

Tabau 4.4 : Extrait abrégé de la liste des noms propres les plus répandus en AS.

ID	Nom	Transcription Phonétique en SAMPA	ID	Nom	Transcription Phonétique en SAMPA
1	إِبْتِسَام	?ibtisa:m	35	حُسَامُ الدِّين	xusa:muddi:n
2	إِدْرِيس	?idri:s	36	حُسَيْن	xusajn
3	أَدَم	a:dam	37	حَفْصَة	xafs'ah
4	أَدْهَم	?adham	38	حَمْرَة	xamzah
5	إِسْمَاعِيل	?isma:?'i:l	39	حَنَان	xana:n
6	العَرَبِي	?al?'arabi:	40	حُورِيَة	xu:rijjah
7	الْهَادِي	?alha:di:	41	حَيَاة	xaja:t
8	الْهُوَارِي	?alhuwa:ri:	42	حَدِيحَة	Xadi:zah
9	إِنْتِشْرَاح	?in ⁰ Sira:x	43	رَبَاب	raba:b
10	إِبْرَاهِيم	?ibra:hi:m	44	رَبِيع	rabi:?'
11	إِسْحَاق	?isxa:q	45	رَبِيعَة	rabi:?'ah
12	إِسْمَاعِيل	?isma:?'i:l	46	رَشِيد	raSi:d
13	إِكْرَام	?ikra:m	47	رَشِيدَة	raSi:dah
14	إِيْمَان	?i:ma:n	48	زَكَرِيَا	zakarija:
15	أَبُو بَكْر	?abu: bakr	49	خَيْرَة	Xajrah
16	أُمُّ كَلْثُوم	?ummu kalTu:m	50	دَاوُد/داوود	da:wud
17	أَمِيرَة	?ami:rah	51	دَلِيلَة	dali:lah
18	أَمِينَة	?ami:nah	52	زَيْد	zajd
19	أَنْس	?anas	53	زَيْدَان	zajda:n
20	أَنْوَر	?anwar	54	زَيْطُوط	zajt'u:t'
21	أَيُّوب	?aju:b	55	سَاكِر	sa:kir
22	أَمَال	a:ma:l	56	سَالِم	sa:lim
23	أَمَنَة	a:minat	57	سَالِمَة	sa:limah
24	بَدْر الدِّين	badra ?addi:n	58	سَالِمِي	sa:lami:
25	بَدْرَان	badra:n	59	سَنَاء	sana:?
26	بَغْدَادِي	baGda:di:	60	سِهَام	siha:m
27	بِلَال	bila:l	61	سُهَيْلَة	suhajlah
28	بَلْقَاسَم	balqa:sam	62	سَيْفُ الدِّين	sajfuddi:ni
29	بَلْقَايِد	balqa:jad	63	سَرَفُ الدِّين	Sarafuddi:ni
30	تَوْفِيق	tawfi:q	64	صَالِح	s'a:lix
31	تَيْجَانِي	ti:za:ni:	65	صَبْرِيَة	s'abrijah
32	جَابِر	za:bar	66	صَلَاحُ الدِّين	s'ala:xuddi:n
33	جَعْفَر	za?'far	67	ضِيَاء	d'iya:?
34	جَمِيلَة	zami:lat	68	طَالِب	t'a:lib

Dans le *Tableau 4.5* nous présentons un aperçu du recueil de textes compilés pour les besoins d'évaluation, avec la transcription phonétique correspondante. Il s'agit du chapitre 67 du Sait Coran, (سورة الملِك) [su:ratulmulki]. Le Saint Coran est un corpus textuel et phonétique très apprécié dans les projets de recherche dans le domaine du TALA (Traitement Automatique de l'AS). Selon Eric Atwell de l'université de Leeds, le Coran est très approprié comme une ressource textuelle et phonétique en raison de sa voyellisation parfaite, de sa disponibilité pour le grand public, sans aucune restriction. De plus, la communauté des experts en Coran est très large et existe depuis plus de 14 siècles, ce qui a permis d'élaborer les normes linguistiques et phonétique. Le grand nombre d'utilisateurs du saint Coran (Lecteurs, récitateurs, auditeurs) assure aux recherches techniques liées à l'AS un bon impact et présente un terrain favorable aux évaluations requises [168].

Tableau 4.5 : Extrait du recueil de textes compilés pour les besoins d'évaluation (su:ratulmulki) avec la transcription phonétique correspondante.

N°	Phrase (Verset)	Transcription Phonétique Manuelle en SAMPA Par des experts Phonéticiens	Transcription Phonétique Automatique par le Système TOP-AS
0	بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ	bismillaxirraxma:nirraxi:m	bismillahirraxma:nirraxi:m
1	تَبَارَكَ الَّذِي بِيَدِهِ الْمُلْكُ وَهُوَ عَلَى كُلِّ شَيْءٍ قَدِيرٌ	taba:rakallaDi: bijadihilmulku wa huwa ?'ala:kulli Saj?'in qadi:r	taba:raka ?'allaDi: bijadihilmulku wa huwa ?'ala:kulli Saj?'in qadi:r
2	الَّذِي خَلَقَ الْمَوْتَ وَالْحَيَاةَ لِيَبْلُوَكُمْ أَيُّكُمْ أَحْسَنُ عَمَلًا وَهُوَ الْعَزِيزُ الْغَفُورُ	?allaDi: Xalalkalmawta walxaja:ta lijabluwakum ?ajjukum ?axsanu ?'amalan wa huwal?'azi:zulGafu:r	?allaDi: Xalaaqalmawta walxaja:ta lijabluwakum ?ajjukum ?axsanu ?'amalan wa huwal?'azi:zulGafu:r
3	الَّذِي خَلَقَ سَبْعَ سَمَاوَاتٍ طِبَاقًا	?allaDi: Xalaqa sab?'a sama:wa:tin t'iba:qa:	?allaDi: Xalaqa sab?'a sama:wa:tin t'iba:qan
"	مَا تَرَى فِي خَلْقِ الرَّحْمَنِ مِنْ تَفَاوُتٍ	ma: tara: fi: Xalqirraxma:ni min'tafa:wutin	ma: tara: fi: Xalqi ?arraxma:ni min'tafa:wutin
"	فَارْجِعِ الْبَصَرَ هَلْ تَرَى مِنْ فُطُورٍ	far3i?'ilbas'ara hal tara: min fut'u:rin	far3i?'ilbas'ara hal tara: min fut'u:rin
4	ثُمَّ ارْجِعِ الْبَصَرَ كَرَّتَيْنِ	Tumar3i?'ilbas'ara karratajn	Tumar3i?'ilbas'ara karratajn
"	يَنْقَلِبُ إِلَيْكَ الْبَصَرُ خَاسِئًا وَهُوَ حَسِيرٌ	Jan'qalib ?ilajkalbas'aru Xa:si?aw'wahuwa xasi:run	jan'qalib ?ilajkalbas'aru Xa:si?aw'wahuwa xasi:run
5	وَلَقَدْ رَئَيْنَا السَّمَاءَ الدُّنْيَا بِمَصَابِيحٍ	wa laqad zaija'nnassama:?addu'nja:bimas'a:bi:xa	Walaqadzaija'nna: ?assama:?a ?addu'nja:bimas'a:bi:xa
"	وَجَعَلْنَاهَا رُجُومًا لِلشَّيَاطِينِ وَاَعْتَدْنَا لَهُمْ عَذَابَ السَّعِيرِ	wa 3a?'alna:ha: ru3umalliSSajati:'i:ni wa?'tadna:lahum ?'aDa:bassa?'ir	wa 3a?'alna:ha: ru3umalliSSajati:'i:ni wa?'tadna:lahum ?'aDa:ba ?assa?'ir
6	وَالَّذِينَ كَفَرُوا بِرَبِّهِمْ عَذَابُ جَهَنَّمَ وَبِئْسَ الْمَصِيرُ	WalillaDi:na kafaru: birabbihim ?'aDa:bu 3axa'nnama wa bi?salmas'ir	Wa lillaDi:na kafaru: birabbihim ?'aDa:bu 3aha'nnama wa bi?sa ?almas'ir
7	إِذَا أُلْفُوا فِيهَا سَمِعُوا لَهَا شَهيقًا وَهِيَ تَفُورُ	?iDa: ?ulqu: fi:ha: sami?'u: laha: Sahi:qaw'wahija tafu:r	?iDa: ?ulqu: fi:ha: sami?'u: laha: Sahi:qaw'wahija tafu:r
8	تَكَادُ تَمَيَّرُ مِنَ الْغَيْظِ	taka:du tamajjazu minalGajD'i	taka:du tamajjazu minalGajD'i
"	كُلَّمَا أَلْقَى فِيهَا فَوْجٌ سَأَلْتَهُمْ خَزَنَتُهَا أَلَمْ يَأْتِكُمْ نَذِيرٌ	kullama: ?ulqija fi:ha: faw3un'sa?alahum Xazanatuha: ?alam ja?tikum naDi:r	kullama: ?ulqija fi:ha: faw3un'sa?alahum Xazanatuha: ?alam ja?tikum naDi:r
9	قَالُوا بَلَى قَدْ جَاءَنَا نَذِيرٌ فَكَذَّبْنَا	qa:lu: bala: qad 3a:?ana: naDi:run'fakaDabna:	qa:lu: bala: qad 3a:?ana: naDi:run'fakaDabna:
"	وَقُلْنَا مَا نَزَّلَ اللَّهُ مِنْ شَيْءٍ إِنْ أَنْتُمْ إِلَّا فِي ضَلَالٍ كَبِيرٍ	Wa qulna: ma: nazzalallahu min'Saj?'in ?an'tu'm ?illa: fi: d'ala:lin kabi:rin	Wa qulna: ma: nazzala ?allahu min'Saj?'in ?an'tu'm ?illa: fi: d'ala:lin kabi:rin
10	وَقَالُوا لَوْ كُنَّا نَسْمَعُ أَوْ نَعْقِلُ مَا كُنَّا فِي أَصْحَابِ السَّعِيرِ	Wa qa:llu: law kunna: nasma?'u ?aw na?'qilu ma:kunna: fi: ?as'xa:bissa?'ir	wa qa:llu: law kunna: nasma?'u ?aw na?'qilu ma:kunna: fi: ?as'xa:bi ?assa?'ir
11	فَاعْتَرَفُوا بِذَنبِهِمْ فَسُحْقًا لِأَصْحَابِ السَّعِيرِ	fa?'tarafu: biDam'bihim fasuxqalli?as'xa:bissa?'iri	fa?'tarafu: biDam'bihim fasuxqalli?as'xa:bi ?assa?'iri
12	إِنَّ الَّذِينَ يَخْشَوْنَ رَبَّهُمْ بِالْغَيْبِ لَهُمْ مَغْفِرَةٌ وَأَجْرٌ كَبِيرٌ	?innallaDi:na jaXSawna rabbahum bilGajbi lahum maGfiratuw'wa?a3run'kabi:run	?inna ?allaDi:na jaXSawna rabbahum bilGajbi lahum maGfiratuw'wa?a3run'kabi:r
13	وَأَسِرُّوا قَوْلَكُمْ أَوِ اجْهَرُوا بِهِ إِنَّهُ عَلِيمٌ بِذَاتِ الصُّدُورِ	wa ?asirru: qawlakum ?awi3xaru: bihi ?innahu ?'ali:mum'biDa:ti:s'udu:ri	wa ?asirru: qawlakum ?awi3haru: bihi ?innahu ?'ali:mum'biDa:ti ?as's'udu:r

14	أَلَا يَعْلَمُ مَنْ خَلَقَ وَهُوَ اللَّطِيفُ الْخَبِيرُ	?ala: ja?lamu man Xalaqa wa huwallati:fulXabi:ru	?ala: ja?lamu man Xalaqa wa huwa ?allat'i:fulXabi:r
15	هُوَ الَّذِي جَعَلَ لَكُمُ الْأَرْضَ دَلُولًا فَامْشُوا فِي مَنَاكِبِهَا	huwallaDi: ?a?ala lakumul?ard'a Dalu:lan famSu: fi: mana:kibiha:	uwa ?allaDi: ?a?ala lakumu ?al?ard'a Dalu:lan famSu: fi: mana:kibiha:
"	وَكُلُوا مِنْ رِزْقِهِ وَإِلَيْهِ النُّشُورُ	wakulu: mirriziqihi wa ?ilajhinnuSu:r	wakulu: mirriziqihi wa ?ilajhi ?annuSu:r
16	ءَأَمِنْتُمْ مَنْ فِي السَّمَاءِ أَنْ يَخْسِفَ بِكُمُ الْأَرْضَ فَإِذَا هِيَ تَمُورُ	?a?amin?tum man fissama?:i ?an jaXsifa bikumul?ard'a fa ?iDa: hija tamu:ru	?a?amin?tum man fi: ?assama?:i ?an jaXsifa bikum ?al?ard'a fa ?iDa: hija tamu:ru
17	أَمْ أَمِنْتُمْ مَنْ فِي السَّمَاءِ أَنْ يُرْسِلَ عَلَيْكُمْ حَاصِبًا فَسَتَعْلَمُونَ كَيْفَ نَذِيرِ	?am ?amin?tum man fissama?:i ?aj?jursila ?alajkum xa:s'iban fasata?lamu:na kajfa naDi:r	?am ?amin?tum man fi : ?assama?:i ?aj?jursila ?alajkum xa:s'iban fasata?lamu:na kajfa naDi:r
18	وَلَقَدْ كَذَّبَ الَّذِينَ مِنْ قَبْلِهِمْ فَكَيْفَ كَانَ نَكِيرِ	Wa laqad kaDDaballaDi:na min?qablihim fakajfa ka:na naki:r	wa laqad kaDDaba ?allaDi:na min?qablihim fakajfa ka:na naki:r
19	أَوَلَمْ يَرَوْا إِلَى الطَّيْرِ فَوْقَهُمْ صَافَاتٍ وَيَقْبِضْنَ مَا يُمَسِّكُهُنَّ إِلَّا الرَّحْمَنُ	?awalam jaraw ?ilattajri fawkahum s'a:fa:tin wa jaqbid'na ma: jumsikuhunna ?illarraxmanu	?awalam jaraw ?ila: ?attayri fawkahum s'a:fa:tin wa jaqbid'na ma: jumsikuhunna ?illarraxman
"	إِنَّهُ بِكُلِّ شَيْءٍ بِصِيرٌ	?innahu bikulli Saj?im?bas'i:r	?innahu bikulli Saj?im?bas'i:r
20	أَمَنْ هَذَا الَّذِي هُوَ جُنْدٌ لَكُمْ	?am?man haDallaDi: huwa ?u?ndullakum	?amman haDa: ?allaDi: huwa ?u?ndullakum
"	يَنْصُرُكُمْ مِنْ دُونِ الرَّحْمَنِ	ja?nsurukum min?du:nirraxmani	ja?nsurukum min?du:ni arraxmani
"	إِنَّ الْكَافِرُونَ إِلَّا فِي غُرُورٍ	?inilka:firu:n ?illa: fi: Guru:rin	?inilka:firu:na ?illa: fi: Guru:r
21	أَمَنْ هَذَا الَّذِي يَزُرُّكُمْ إِنْ أَمْسَكَ رِزْقَهُ	?am?man ha:DallaDi: jarzuqukum ?in ?amsaqa rizkahu	?amman ha:Da:llaDi: jarzuqukum ?in ?amsaka rizkahu
"	بَلْ لَجُوا فِي عُتُوٍّ وَنُفُورٍ	balla?zu: fi: ?utuwwiw?wanufu:rin	bal la?zu: fi: ?utuwwiww?anufu:r
22	أَفَمَنْ يَمْشِي مُكِبًّا عَلَى وَجْهِهِ أَهْدَى	?afamaj?jamSi: mukibban ?ala: wa?hihi ?ahda:	?afamaj?jamSi: mukibban ?ala: wa?hihi ?ahda:
"	أَمَنْ يَمْشِي سَوِيًّا عَلَى صِرَاطٍ مُسْتَقِيمٍ	?am?maj?jamSi: sawijjan ?ala: s'ira:t'im?mustaqi:m	?am?maj?jamSi: sawijjan ?ala: s'ira:t'im?mustaqi:m
23	قُلْ هُوَ الَّذِي أَنْشَأَكُمْ	qul huwallaDi: ?a?nSa?akum	qul huwa ?allaDi: ?a?nSa?akum
"	وَجَعَلَ لَكُمْ السَّمْعَ وَالْأَبْصَارَ وَالْأَفْئِدَةَ قَلِيلًا مَّا تَشْكُرُونَ	wa?a?ala lakumussam?a wal?abs'a:ra wal?af?idata qali:lam?ma: taSkuru:n	wa?a?ala lakumu ?assam?a wal ?absa:ra wal ?af?idata qali:lam?ma: taSkuru:n
24	قُلْ هُوَ الَّذِي ذَرَأَكُمْ فِي الْأَرْضِ وَإِلَيْهِ تُحْشَرُونَ	qul huwallaDi: Dara?akum fil?ard'i wa ?ilajhi tuxSaru:n	qul huwa allaDi: Dara?akum fi:l ?ard'i wa ?ilajhi tuxSaru:n
25	وَيَقُولُونَ مَتَى هَذَا الْوَعْدُ إِنْ كُنْتُمْ صَادِقِينَ	wajaqu:lu:na mata: haDalwa?du ?in kun?tu?m s'a:diqin:	wajaqu:l:na mata: ha:Dalwa?du ?in kun?tu?m s'a:diqi:n
26	قُلْ إِنَّمَا الْعِلْمُ عِنْدَ اللَّهِ وَإِنَّمَا أَنَا نَذِيرٌ مُبِينٌ	qul ?innamal?ilmu ?indallahi wa ?innama: ?ana: naDi:rum?mubi:nun	qul ?innamal?ilmu ?inda alla:hi wa ?innama: ?ana: naDi:rum?mubi:nun
27	فَلَمَّا رَأَوْهُ زُلْفَةً سِيئَتْ وُجُوهُ الَّذِينَ كَفَرُوا	falamma: ra?awhu zulfatan si:?at wu?u:hullaDi:na kafaru:	falamma: ra?awhu zulfatan?si:?at wu?u:hu ?allaDi:na kafaru:
"	وَقِيلَ هَذَا الَّذِي كُنْتُمْ بِهِ تَدَّعُونَ	waqi:la ha:DallaDi: kun?tu?m bihi tadda?u:n	waqi:la haDa: ?allaDi: kun?tu?m bihi tadda?u:n
28	قُلْ أَرَأَيْتُمْ إِنْ أَهْلَكَنِیَ اللَّهُ وَمَنْ مَعِيَ أَوْ رَحِمَنَا	qul ?ara?ajtum ?in ?ahlakanijallahu wa mam?ma?ija ?aw raximana:	qul ?ara?ajtum ?in ?ahlakanija ?allahu wa mam?ma?ija ?aw raximana:
"	فَمَنْ يُجِيرُ الْكَافِرِينَ مِنْ عَذَابِ أَلِيمٍ	famaj?ju?i:rulka:firi:na min ?aDa:bin ?ali:m	famaj?ju?i:rulka:firi:na min?aDa:bin ?ali:m
29	قُلْ هُوَ الرَّحْمَنُ أَمَّا بِهِ وَعَلَيْهِ تَوَكَّلْنَا فَسَتَعْلَمُونَ مَنْ هُوَ فِي ضَلَالٍ مُبِينٍ	qul huwarraxmanu ?amanna: bihi wa ?alajhi tawakkalna: fasata?lamu:na man huwa fi: D'al:lim?mubi:n	qul huwarraxmanu ?a:manna: bihi wa ?alajhi tawakkalna: fasata?lamu:na man huwa fi: D'ala:lim?mubi:n
30	قُلْ أَرَأَيْتُمْ إِنْ أَصْبَحَ مَاؤُكُمْ غَوْرًا فَمَنْ يَأْتِيكُمْ بِمَاءٍ مَعِينٍ.	qul ?ara?ajtum ?in ?asbaxa ma:?ukum Gawran?famaj?ja:ti:kum bima:?im?ma?i:n	qul ?ara?ajtum ?in ?asbaxa ma:?ukum Gawran?famaj?ja:ti:kum bima:?im?ma?i:n
	صَدَقَ اللَّهُ الْعَظِيمُ	S'adaqallahul?aD'i:m	s'adaqallahul?aD'i:m

Le tableau 4.6 présente un aperçu sur une autre partie de notre compilation de textes pour l'évaluation du système TOP-AS. Il s'agit de quelques recueils de poèmes de l'ère préislamique (الشعر الجاهلي) avec leurs transcriptions de référence, comparées aux transcriptions automatiques du système TOP-AS.

Tableau 4.6 : Extrait du recueil de textes compilés pour les besoins d'évaluation (Quelques poèmes de l'ère pré-islamique) avec la transcription phonétique correspondante.

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
عبيد بن الأبرص Obaïd ben al-Abras		
يَا مَنْ لَبِزِي أَيْبْتُ اللَّيْلَ أَزُقُبُهُ مِنْ عَارِضٍ كَنِيَاضِ الصُّبْحِ لَمَّاحِ	ja: mallibarqin ?abi:tullajla ?arqubuhu min ?'a:rid'in kabaja:d'is's'ubxi lamma:xi	ja: mallibarqin ?abi:tullajla ?arqubuhu min ?'a:rid'in kabaja:d'is's'ubxi lamma:xi
ذَانِ مُسِفِّ فُوقِ الْأَرْضِ هَيْدَبُهُ يَكَادُ يَدْفَعُهُ مَنْ قَامَ بِالرَّاحِ	da:nim ⁿ musiffin fuwajqal?ard'i hajdabuhu jaka:du jadfa?'uhu man qa:ma birra:xi	da:nim ⁿ musiffin fuwajqal?ard'i hajdabuhu jaka:du jadfa?'uhu man qa:ma birra:xi
فَمَنْ بِنَجْوَتِهِ كَمَنْ بِمَحْفَلِهِ وَالْمُسْتَكِنُ كَمَنْ يَمِثِّي بِقِرْوَاكِ	famam ⁿ bina3watihi kaman bimaxfalihi walmustakinnu kamaj ⁿ jamSi: biqirwa:xi	famam ⁿ bina3watihi kaman bimaxfalihi walmustakinnu kamaj ⁿ jamSi: biqirwa:xi
كَأَنَّ رَيْقَهُ لَمَّا عَلَا شَطْبًا أَقْرَابُ أْبْلَقِي يَنْفِي الْخَيْلَ رَمَّاحِ	ka?anna rajjiqahu lamma: ?'ala: Sat'iban ?aqra:bu ?ablaqij ⁿ janfilXajla ramma:xi	ka?anna rajjiqahu lamma: ?'ala: Sat'iban ?aqra:bu ?ablaqij ⁿ janfilXajla ramma:xi
فَالْتَجَّ أَعْلَاهُ ثُمَّ ارْتَجَّ أَسْفَلُهُ وَصَاقَ ذَرْعًا بِحَمَلِ الْمَاءِ مُنْصَاحِ	falta33a ?a?'la:hu Tummarta33a ?asfalu wad'a:qa Dar?'am ⁿ bixamliima:?'i mu ⁿ ns'a:xi	falta33a ?a?'la:hu Tummarta33a ?asfalu wad'a:qa Dar?'am ⁿ bixamliima:?'i mu ⁿ ns'a:xi
كَأَنَّمَا بَيْنَ أَعْلَاهُ وَأَسْفَلِهِ رَيْطٌ مُنْشَرَّةٌ أَوْ ضَوْءٌ مُصْبَاحِ	ka?annama: bajna ?a?'la:hu wa?asfalihi rajt'um ⁿ munaSSaratun ?aw d'aw?u mis'ba:xi	ka?annama: bajna ?a?'la:hu wa?asfalihi rajt'um ⁿ munaSSaratun ?aw d'aw?u mis'ba:xi
كَأَنَّ فِيهِ عِشَارًا جِلَّةً شُرْفًا شُعْنًا لِهَامِيمٍ قَدْ هَمَّتْ بِإِشْرَاحِ	ka?anna fi:hi ?'iSa:ran zillatan Surufan Su?'Tallaha:mi:ma qad hammat bi?'irSa:xi	ka?anna fi:hi ?'iSa:ran zillatan Surufan Su?'Tallaha:mi:ma qad hammat bi?'irSa:xi
بُخَّا حَنَاجِرُهَا هُدْلًا مَشَافِرُهَا تُسِيمٌ أَوْلَادَهَا فِي قَرْقَرٍ ضَاحِي	buxxan xana:ziruha: hudlam ⁿ maSa:firuha: tusi:mu ?awla:daha: fi: qarqarin d'a:xi:	buxxan xana:ziruha: hudlam ⁿ maSa:firuha: tusi:mu ?awla:daha: fi: qarqarin d'a:xi:
هَبَّتْ جَنُوبٌ بِأَوْلَادِهِ وَمَالَ بِهِ أَعْجَازُ مُزْنٍ يَسُحُّ الْمَاءَ دَلَّاحِ	habbat zanu:bum ⁿ bi?u:la:hu wama:la bihi ?a?'za:zu muznij ⁿ jasuxxulma:?'a dala:axi	habbat zanu:bum ⁿ bi?u:la:hu wama:la bihi ?a?'za:zu muznij ⁿ jasuxxulma:?'a dala:axi
فَأَصْبَحَ الرَّوْضُ وَالْقَيْعَانُ مُمْرِعَةً مِنْ بَيْنِ مُرْتَفِقٍ فِيهِ وَمُنْطَاحِ	fa?as'baxarrawd'u wa?alqi:?'a:nu mumri?'atan mim ⁿ bajni murtafiqin fi:hi wamu ⁿ nt'a:xi	fa?as'baxarrawd'u wa?alqi:?'a:nu mumri?'atan mim ⁿ bajni murtafiqin fi:hi wamu ⁿ nt'a:xi

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
عدي بن زيد Ouday ben Zaid		
أَرِقْتُ لِمُكْفَهِّرِ بَاتَ فِيهِ بَوَارِقُ يَرْتَقِينَ رُؤُوسَ شَيْبِ	?ariqtu limukfahirrim ⁿ ba:ta fi:hi bawa:riqu jartaqi:na ru?u:sa Si:bi	?ariqtu limukfahirrim ⁿ ba:ta fi:hi bawa:riqu jartaqi:na ru?u:sa Si:bi
تَلُوحُ الْمَشْرِفِيَّةُ فِي ذُرَاهُ وَيَجْلُو صَفْحَ دَخْدَارٍ قَشِيبِ	talu:xulmaSrafijjatu fi: Dura:hu waja3lu: s'afxaxdaXda:rin qaSi:bi	talu:xulmaSrafijjatu fi: Dura:hu waja3lu: s'afxaxdaXda:rin qaSi:bi
كَأَنَّ مَاتِمًا بَاتَتْ عَلَيْهِ خَضْبَنَ مَالِيًا بِدَمِ صَبِيبِ	ka?anna ma?a:timam ⁿ ba:tat ?'alajhi Xad'abna ma?a:lijam ⁿ bidamin s'abijbi	ka?anna ma?a:timam ⁿ ba:tat ?'alajhi Xad'abna ma?a:lijam ⁿ bidamin s'abijbi
يُلَالِئُنَّ الْأَكْفَ عَلَى عَدِيٍّ وَيُعْطَفُ رَجْعُهُنَّ إِلَى الْجُيُوبِ	jula?li?nal?akuffa ?'ala: ?'adijjin waju?'t'afu raz?'uhunna ?ilal3uju:bi	jula?li?nal?akuffa ?'ala: ?'adijjin waju?'t'afu raz?'uhunna ?ilal3uju:bi
سَقَى بَطْنَ الْعَمِيقِ إِلَى أَفَاقِ فَفَأُثُورٍ إِلَى لَبِّبِ الْكَثِيبِ	saqa: bat'na ?al?'aqi:qi ?ila: ?ufa:qin fafa:Tu:rin ?ila: lababilkaTi:bi	saqa: bat'na ?al?'aqi:qi ?ila: ?ufa:qin fafa:Tu:rin ?ila: lababilkaTi:bi
فَرَوَى قُلَّةَ الْأَدْحَالِ وَنَلَّ فَمَلَجًا فَالْنَبِيِّ فَذَا كَرِيبِ	farawwa: qullatal?adxa:li wablun fafal3an fannabijja faDa: kari:bi	farawwa: qullatal?adxa:li wablun fafal3an fannabijja faDa: kari:bi
فَمِنْهُ دِيمَةٌ وَطَفَاءٌ سَكَبُ وَدُو نَزَلِ تَفَرَّعُ فِي السُّيُوبِ	faminhu di:matuw ⁿ wat'fa:?u sakbun waDu: nazalin tafarra?'u fissuju:bi	faminhu di:matuw ⁿ wat'fa:?u sakbun waDu: nazalin tafarra?'u fissuju:bi
كَأَنَّ دُفُوقَ جَوْنٍ تَعْتَرِيهِ تُجَانِبُ قَاصِبًا فَحَنِينِ نَيْبِ	ka?anna dufu:qa zawnin ta?'tari:hi tu3a:nibu qa:s'iban faxani:na ni:bi	ka?anna dufu:qa zawnin ta?'tari:hi tu3a:nibu qa:s'iban faxani:na ni:bi
يَجِيءُ بِمَا أَمَدَّتْهُ الثُّرَيَّا مُعِيرًا أَمْرَهُ دَرَرَ الْجَنُوبِ	ja3i:?u bima: ?amaddathu ?aTTurajja: mu?'i:ran ?amrahu diraral3anu:bi	ja3i:?u bima: ?amaddathu ?aTTurajja: mu?'i:ran ?amrahu diraral3anu:bi

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
الأعشى Al-Archa		
يَا مَنْ يَرَى عَارِضًا قَدِ بَتُّ أَرْقُبُهُ كَأَنَّهَا الْبَرْقُ فِي حَافَاتِهِ الشُّعْلُ	ja: maj ⁿ jara: ?'a:rid'an qad bittu ?arqubuhu ka?annama: ?albarqu fi: xa:fa:tihISSu?'alu	ja: maj ⁿ jara: ?'a:rid'an qad bittu ?arqubuhu ka?annama: ?albarqu fi: xa:fa:tihISSu?'alu
لَهُ رِدَافٌ وَجَوْرٌ مُفَامٌ عَمِلُ مُنَطَّقٌ بِسَجَالِ الْمَاءِ مُتَّصِلُ	lahu rida:fuw ⁿ wazawzum ⁿ muf?amun ?'amilun munat't'aqum ⁿ bisiza:lilma:?'i muttas'ilu	lahu rida:fuw ⁿ wazawzum ⁿ muf?amun ?'amilun munat't'aqum ⁿ bisiza:lilma:?'i muttas'ilu
لَمْ يُلْبِي اللَّهَ عَنْهُ حِينَ أَرْقُبُهُ وَلَا اللَّذَاذَةَ مِنْ كَأْسٍ وَلَا الْكَسَلَ	lam julhinillahwu ?'a ⁿ nhu xi:na ?arqubuhu walallaDa:Datu min ⁿ ka?siw ⁿ walalkasalu	lam julhinillahwu ?'a ⁿ nhu xi:na ?arqubuhu walallaDa:Datu min ⁿ ka?siw ⁿ walalkasalu
فَقُلْتُ لِلشَّرْبِ فِي دُرْنِي وَقَدْ تَمَلُّوا شِيمُوا، وَكَيْفَ يَشِيمُ الشَّارِبُ الْتَّمِلُ	faqultu liSSarbi fi: durna: waqad Tamilu: Si:mu: wakajfa jaSi:muSSa:ribuTTamilu	faqultu liSSarbi fi: durna: waqad Tamilu: Si:mu: wakajfa jaSi:muSSa:ribuTTamilu
بَرْقًا يُضِيءُ عَلَى الْأَجْرَاعِ مَسْقَطُهُ وَبِالْحَيْيَةِ مِنْهُ عَارِضٌ هَاطِلُ	barqaj ⁿ jud'i:?'u ?'alal?a3za:?'i masqat'uhu wabilXabijjati minhu ?'a:rid'un hat'ilu	barqaj ⁿ jud'i:?'u ?'alal?a3za:?'i masqat'uhu wabilXabijjati minhu ?'a:rid'un hat'ilu
قَالُوا نِمَارًا فَبَطْنُ الْخَالِ جَادَهُمَا فَالْعَسَجِدِيَّةُ فَالْأَبْلَاءُ فَالرَّجُلُ	qa?alu: nima:run fabat'nulXa:li za:dahuma: fa?al?'as3adijjatu fal?abla:?'u farrizalu	qa?alu: nima:run fabat'nulXa:li za:dahuma: fa?al?'as3adijjatu fal?abla:?'u farrizalu
فَالسَّفْحُ يَجْرِي فَخَزِيرٌ فَبَرْقَتُهُ حَتَّى تَدَافِعَ الرَّبُّو فَالْجَبَلُ	fa?assafxu jazri: faXinzi:run faburqatuhu xatta: tada:fa?'a ?arrabwu fa?al3abalu	fa?assafxu jazri: faXinzi:run faburqatuhu xatta: tada:fa?'a ?arrabwu fa?al3abalu
حَتَّى تَحْمَلَ مِنْهُ الْمَاءَ تَكْلِفَةً رَوْضُ الْقَطَا فَكَنْيَبُ الْغَيْنَةِ السَّهْلُ	xatta: taxammala minhulma:?'a taklifatan rawd'ulqat'a: fakaTi:bulGajnati ?assahilu	xatta: taxammala minhulma:?'a taklifatan rawd'ulqat'a: fakaTi:bulGajnati ?assahilu
يَسْقِي دِيَارًا لَهَا قَدْ أَصْبَحَتْ عُرْبًا زُورًا تَجَانَفَ عَنْهَا الْقَوْدُ وَالرَّسْلُ	jasqi: dija:rallaha: qad ?as'baxat ?'uzuban zu:ran tazza:nafa ?'a ⁿ nha: ?alqawdu warrasalu	jasqi: dija:rallaha: qad ?as'baxat ?'uzuban zu:ran tazza:nafa ?'a ⁿ nha: ?alqawdu warrasalu

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
أوس بن حجر Aws ben Hajar		
إِنِّي أَرَقْتُ وَلَمْ تَأْرُقْ مَعِيَ صَاحِي لِمُسْتَكْفٍ بُعِيدِ النَّوْمِ لَوَاحِ	?inni: ?ariqtu walam ta?raq ma?i: s'a:xi: limustakfim ⁿ bu?ajdannawmi lawwa:xi	?inni: ?ariqtu walam ta?raq ma?i: s'a:xi: limustakfim ⁿ bu?ajdannawmi lawwa:xi
قَدْ نِمْتَ عَنِّي وَبَاتَ الْبَرْقُ يُسْهِرُنِي كَمَا اسْتَضَاءَ يَهُودِيٌّ بِمِصْبَاحِ	qad nimta ?'anni: waba:talbarqu jushiruni: kamastad'a: ?a jahu: dijjum ⁿ bimis'ba:xi	qad nimta ?'anni: waba:talbarqu jushiruni: kamastad'a: ?a jahu: dijjum ⁿ bimis'ba:xi
يَا مَنْ لِبَرْقِ أَيْبُتِ اللَّيْلِ أَرْقُبُهُ فِي عَارِضِ كَمْضِيءِ الصُّبْحِ لَمَّاحِ	ja: mallibarqin ?abi:tullajla ?arqubuhu fi: ?'a:rid'in kamud'i: ?is's'ubxi lamma:xi	ja: mallibarqin ?abi:tullajla ?arqubuhu fi: ?'a:rid'in kamud'i: ?is's'ubxi lamma:xi
ذَانِ مُسْفٍ فَوْقَ الْأَرْضِ هَيْدَبُهُ يَكَادُ يَدْفَعُهُ مَنْ قَامَ بِالرَّاحِ	da:nim ⁿ musiffin fuwajqal?ard'i hajdabuhu jaka:du jadfa?'uhu man qa:ma birra:xi	da:nim ⁿ musiffin fuwajqal?ard'i hajdabuhu jaka:du jadfa?'uhu man qa:ma birra:xi
كَأَنَّ رَيْقَهُ لَمَّا عَلَا شَطْبًا أَقْرَابُ أْبْلَقِ يَنْفِي الْخَيْلَ رَمَّاحِ	ka?anna rajjiqahu lamma: ?'ala: Sat'iban ?aqra:bu ?ablaqa janfilXajla ramma:xi	ka?anna rajjiqahu lamma: ?'ala: Sat'iban ?aqra:bu ?ablaqa janfilXajla ramma:xi
هَبَّتْ جُنُوبٌ بِأَعْلَاهُ وَمَالَ بِهِ أَعْجَازُ مَزْنٍ يَسُحُّ الْمَاءَ دَلَّاحِ	habbat zanu:bum ⁿ bi?a?'la:hu wama:la bihi ?a?'za:zu muznij ⁿ jasuxxulma: ?a dala:axi	habbat zanu:bum ⁿ bi?a?'la:hu wama:la bihi ?a?'za:zu muznij ⁿ jasuxxulma: ?a dala:axi
فَالْتَجَّ أَعْلَاهُ ثُمَّ ارْتَجَّ أَسْفَلُهُ وَضَاقَ دَرْعًا بِحَمَلِ الْمَاءِ مُنْصَاحِ	falta33a ?a?'la:hu Tummarta33a ?asfaluhu wad'a:qa Dar?'am ⁿ bixamlilma: ?i mu ⁿ ns'a:xi	falta33a ?a?'la:hu Tummarta33a ?asfaluhu wad'a:qa Dar?'am ⁿ bixamlilma: ?i mu ⁿ ns'a:xi
كَأَنَّمَا بَيْنَ أَعْلَاهُ وَأَسْفَلِهِ رَيْطٌ مُنْشَرَّةٌ أَوْ ضَوْءٌ مُصْبَاحِ	ka?annama: bajna ?a?'la:hu wa?asfalihi rajt'um ⁿ munaSSaratun ?aw d'aw?u mis'ba:xi	ka?annama: bajna ?a?'la:hu wa?asfalihi rajt'um ⁿ munaSSaratun ?aw d'aw?u mis'ba:xi
يُزْعُ جِلْدَ الْحَصَى أَجْسَ مُبْتَرِكٌ كَأَنَّهُ فَاحِصٌ أَوْ لَاعِبٌ دَاجِي	janza?'u zildalxas'a: ?a3aSSu mubtarikun ka?annahu fa:xis'un ?aw la:?'ibun da:xi:	janza?'u zildalxas'a: ?a3aSSu mubtarikun ka?annahu fa:xis'un ?aw la:?'ibun da:xi:
فَمَنْ بِنَجْوَتِهِ كَمَنْ بِمَحْفَلِهِ وَالْمُسْتَكِنُ كَمَنْ يَمِثِّي بِقِرْوَاكِ	famam ⁿ bina3watihi kamam ⁿ bimaxfalihi walmustakinnu kamaj ⁿ jamSi: biqirwa:xi	famam ⁿ bina3watihi kamam ⁿ bimaxfalihi walmustakinnu kamaj ⁿ jamSi: biqirwa:xi
كَأَنَّ فِيهِ عِشَارًا جِلَّةً شُرْفًا شُعْنًا لَهَا مِيمٌ قَدْ هَمَّتْ بِإِشْرَاحِ	ka?anna fi:hi ?'iSa:ran zillatan Surufan Su?'Tallaha:mi:ma qad hammat bi?'irSa:xi	ka?anna fi:hi ?'iSa:ran zillatan Surufan Su?'Tallaha:mi:ma qad hammat bi?'irSa:xi
هُدْلًا مَشَافِرُهَا بَحًّا حَنَاجِرُهَا تُرْجِي مَرَابِعَهَا فِي صَحْصَحِ صَاحِي	hudlam ⁿ maSa:firuha: buxxan xana:ziruha: tuzzi: mara:bi:?'aha: fi: s'axs'axin d'a:xi:	hudlam ⁿ maSa:firuha: buxxan xana:ziruha: tuzzi: mara:bi:?'aha: fi: s'axs'axin d'a:xi:
فَأَصْبَحَ الرَّوْضُ وَالْقَيْعَانُ مُمْرِعَةً مَنْ بَيْنَ مُرْتَفِقِي مِثْمَا وَمُنْطَاحِ	fa?as'baxarrawd'u wa?alqi:?'a:nu mumri?'atan mim ⁿ bajni murtafiqim ⁿ minha: wamu ⁿ nt'a:xi	fa?as'baxarrawd'u wa?alqi:?'a:nu mumri?'atan mim ⁿ bajni murtafiqim ⁿ minha: wamu ⁿ nt'a:xi
سَقَى دِيَارَ بَنِي عَوْفٍ وَسَاكِنَتَهَا وَدَارَ عَلَقَمَةَ الْخَيْرِ بِنِ صَبَّاحِ	saqa: dija:ra bani: ?'awfiw ⁿ wasa:kinaha: wada:ra ?'alqamatalXajri bni s'abba:xi	saqa: dija:ra bani: ?'awfiw ⁿ wasa:kinaha: wada:ra ?'alqamatalXajri bni s'abba:xi

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
امرؤ القيس Imru' al-Qays		
أَحَارِ تَرَى بَرْقًا كَأَنَّ وَمِيضَهُ كَلَمَعَ الْيَدَيْنِ فِي حَبِيٍّ مُكَلَّلٍ	?axa:ri tara: barqan ka?anna wami:d'ahu kalam?'iljadajni fi: xabijjim ⁿ mukallali	?axa:ri tara: barqan ka?anna wami:d'ahu kalam?'iljadajni fi: xabijjim ⁿ mukallali
يُضِيءُ سَنَاهُ أَوْ مَصَابِيحُ رَاهِبٍ أَهَانَ السَّلِيطَ فِي الدُّبَالِ الْمُفْتَلِّ	jud'i:?u sana:hu ?aw mas'a:bi:xu ra:hibin ?aha:nassali:t'a fiDDuba:lilmufattali	jud'i:?u sana:hu ?aw mas'a:bi:xu ra:hibin ?aha:nassali:t'a fiDDuba:lilmufattali
فَعَدْتُ لَهُ وَصُحْبَتِي بَيْنَ حَامِرٍ وَبَيْنَ إِكَامٍ بَعْدَ مَا مُتَأَمَّلٍ	qa?'adtu lahu was'uxbati: bajna xa:mirin wabajna ?ika:mim ⁿ 'bu?'da ma: muta?ammali	qa?'adtu lahu was'uxbati: bajna xa:mirin wabajna ?ika:mim ⁿ 'bu?'da ma: muta?ammali
وَأَضْحَى يَسُحُ الْمَاءَ عَن كُلِّ فَيْقَةٍ يَكُبُّ عَلَى الْأَذْقَانِ دَوْحَ الْكَنْهَبِلِ	wa?ad'xa: jasuxxulma:?a ?'an kulli fi:qatin jakubbu ?'ala: ?al?aDqa:ni dawxalkanahbali	wa?ad'xa: jasuxxulma:?a ?'an kulli fi:qatin jakubbu ?'ala: ?al?aDqa:ni dawxalkanahbali
وَتَيْمَاءٌ لَمْ يَتْرُكْ بِهَا جِدْعَ نَخْلَةٍ وَلَا أُطْمًا إِلَّا مَشِيدًا بِجَنْدَلِ	watajma:?a lam jatruk biha: 3iD?'a naXlatin wala: ?ut'uman ?lla: maSi:d ⁿ bi3andali	watajma:?a lam jatruk biha: 3iD?'a naXlatin wala: ?ut'uman ?lla: maSi:d ⁿ bi3andali
كَأَنَّ طَمِيَّةَ الْمُجَيْمِرِ غُدُوَّةٌ مِنَ السَّيْلِ وَالْغُثَاءِ فَلَكَّةُ مِغْزَلِ	ka?anna t'amijjatalmu3ajmiri Gudwatan minassajli walGuTTa:?i falkatu miGzali	ka?anna t'amijjatalmu3ajmiri Gudwatan minassajli walGuTTa:?i falkatu miGzali
كَأَنَّ أَبَانًا فِي أَفَانِينَ وَدَقِيهِ كَبِيرٌ أَنَاسٍ فِي بَجَادٍ مُرْمَلِ	ka?anna ?aba:nan fi: ?afa:ni:na wadqih kabi:ru ?una:sin fi: bi3a:dim ⁿ muzammali	ka?anna ?aba:nan fi: ?afa:ni:na wadqih kabi:ru ?una:sin fi: bi3a:dim ⁿ muzammali
وَأَلْقَى بِصَحْرَاءِ الْعَبِيطِ بَعَاغُهُ نُزُولَ الْيَمَانِيِّ ذِي الْعِيَابِ الْمُخَوَّلِ	wa?alqa: bis'axra: ?ilGabi:t'i ba?'a:?'ahu nuzu:laljama:ni: Dil?'ija:bi ?almuXawwali	wa?alqa: bis'axra: ?ilGabi:t'i ba?'a:?'ahu nuzu:laljama:ni: Dil?'ija:bi ?almuXawwali
كَأَنَّ سِبَاعًا فِيهِ غَرْقَى غُدِيَّةٌ بِأَرْجَائِهِ الْقُصُوى أَنَابِيشُ عُنْصُلِ	ka?anna siba:?'an fi:hi Garqa: Gudajjatan bi?ar3a: ?ihi ?alqus'wa: ?ana:bi:Su ?'u ⁿ ns'uli	ka?anna siba:?'an fi:hi Garqa: Gudajjatan bi?ar3a: ?ihi ?alqus'wa: ?ana:bi:Su ?'u ⁿ ns'uli
عَلَى قَطْنٍ بِالسَّيْمِ أَيْمَنُ صَوْبِهِ وَأَيْسَرُهُ عَلَى السِّتَارِ فَيَذْبُلِ	?'ala: qat'anim ⁿ biSSajmi ?ajmanu s'awbihi wa?ajsaruhu ?'alassita:ri fajaDbuli	?'ala: qat'anim ⁿ biSSajmi ?ajmanu s'awbihi wa?ajsaruhu ?'alassita:ri fajaDbuli
وَأَلْقَى بِبُسَيَانَ مَعَ اللَّيْلِ بَرْكَهُ فَأَنْزَلَ مِنْهُ الْعُصْمَ مِنْ كُلِّ مَنْزِلِ	wa?alqa: bibusja:nim ⁿ ma?'allajli barkahu fa?anzala mi ⁿ nhul?'us'ma min ⁰ kulli manzili	wa?alqa: bibusja:nim ⁿ ma?'allajli barkahu fa?anzala mi ⁿ nhul?'us'ma min ⁰ kulli manzili

Vers du poème	Transcription Phonétique SAMPA de Référence	Transcription Automatique par le système TOP-AS
ربيعة Labeed bin Rabi'ah		
يَا هَلْ تَرَى الْبَرْقَ بَتُّ أَرْقُبُهُ يُرْجِي حَبِيًّا إِذَا حَبَا نَقَبَا	ja: hal taralbarqa bittu ?arqubuhu juzzi: xabijjan ?iDa: Xaba: Taqaba:	ja: hal taralbarqa bittu ?arqubuhu juzzi: xabijjan ?iDa: Xaba: Taqaba:
فَعَدْتُ وَحْدِي لَهُ، وَقَالَ أَبُو لَيْلَى: مَتَى يَغْتَمِنُ فَقَدْ دَابَا	qa?'adtu waxdi: lahu wqa:la ?abu: ljla: mata: jaGtamin faqad da?aba:	qa?'adtu waxdi: lahu wqa:la ?abu: ljla: mata: jaGtamin faqad da?aba:
كَأَنَّ فِيهِ لَمَّا ارْتَفَقْتُ لَهُ رَيْطًا وَمِرْبَاعَ غَانِمٍ لَجَبَا	ka?anna fi:hi lammartafaqtu lahu rajt'aw"wamirba:?'a Ga:nimillazaba:	ka?anna fi:hi lammartafaqtu lahu rajt'aw"wamirba:?'a Ga:nimillazaba:
فَجَادَ رَهْوًا إِلَى مَدَاخِلِ فَالِصُّ صُحْرَةَ أَمْسَتْ نِجَاجُهُ عَصَبَا	faʒa:da rahwan ?ila: mada:Xila fas' s'uxrati ?amsat ni?'a:ʒuhu ?'us'aba:	faʒa:da rahwan ?ila: mada:Xila fas' s'uxrati ?amsat ni?'a:ʒuhu ?'us'aba:
فَحَدَّرَ الْعُصْمَ مِنْ عَمَايَةَ لِسْدِ سَهْلٍ وَقَضَى بِصَاحَةِ الْأَرْبَا	faxaddaral?'us'ma min ?'ama:jata lis sahli waqad'd'a: bis'a:xata ?al?araba:	faxaddaral?'us'ma min ?'ama:jata lis sahli waqad'd'a: bis'a:xata ?al?araba:
فَالْمَاءُ يَجْلُو مُتَوَهِّنَ كَمَا يَجْلُو التَّلَامِيذُ لَوْلَا قَسِبَا	falma:?'u jaʒlu: mutu:nahunna kama: jaʒluttala:mi:Du lu?ulu?uan qaSiba:	falma:?'u jaʒlu: mutu:nahunna kama: jaʒluttala:mi:Du lu?ulu?uan qaSiba:
لَأَقَى الْبَدِيِّ الْكِلَابَ فَاعْتَلَجَا مَوْجَ أَيْبِهِمَا مِمَّنْ غَلَبَا	la:aqalbadijju ?alkila:ba fa?'talaʒa: mawʒu ?atijjahhima: liman Galaba:	la:aqalbadijju ?alkila:ba fa?'talaʒa: mawʒu ?atijjahhima: liman Galaba:
فَدَعَدَا سُرَّةَ الرِّكَاةِ كَمَا دَعَدَعَ سَاقِي الْأَعَاجِمِ الْغَرَبَا	fada?'da?'a: surratarraka:?'i kama: da?'da?'a sa:qil?a?'a:ʒimi ?alGaraba:	fada?'da?'a: surratarraka:?'i kama: da?'da?'a sa:qil?a?'a:ʒimi ?alGaraba:
فَكُلُّ وَادٍ هَدَّتْ حَوَالِبُهُ يَقْدِفُ خُضْرَ الدَّبَائِ فَالْخُشْبَا	fakullu wa:din haddat xawa:libuhu jaqDifu Xud'raddaba:?'i fa?alXuSuba:	fakullu wa:din haddat xawa:libuhu jaqDifu Xud'raddaba:?'i fa?alXuSuba:
مَالَتْ بِهِ نَحْوَهَا الْجَنُوبُ مَعَا ثُمَّ ازْدَهَتْهُ السَّمَالُ فَاثْقَلَبَا	ma:lat bihi naxwaha: ?alʒanu:bu ma?'an Tummazdahathu ?aSSama:lu fanqalaba:	ma:lat bihi naxwaha: ?alʒanu:bu ma?'an Tummazdahathu ?aSSama:lu fanqalaba:
فَقُلْتُ صَابَ الْأَعْرَاضَ رَيْفُهُ يَسْقِي بِلَادًا قَدْ أَمَحَلَتْ حِقَبَا	faqultu s'a:bal?a?'ra:d'a rajjiquhu jasqi: bila:dan qad ?amxalat xiqaba:	faqultu s'a:bal?a?'ra:d'a rajjiquhu jasqi: bila:dan qad ?amxalat xiqaba:
لِتَرَعَ مِنْ نَبْتِهِ أُسَيْمٌ إِذَا أَنْبَتَ حُرُّ الْبُقُولِ وَالْعُشْبَا	litar?'a minnabtihī ?usajmu ?iDa: ?am ⁿ bata xurralbuqu:li wal?'uSuba:	litar?'a minnabtihī ?usajmu ?iDa: ?am ⁿ bata xurralbuqu:li wal?'uSuba:
وَلَيَرَعُهُ قَوْمُهَا فَإِنَّهُمْ مِنْ خَيْرِ حَيِّ عِلْمَتُهُمْ حَسَبَا	waljar?'ahu qawmuha: fa?innahumu min Xajri xajjin ?'alimtuhum xasaba:	waljar?'ahu qawmuha: fa?innahumu min Xajri xajjin ?'alimtuhum xasaba:

4.3.2 Evaluation

4.3.2.1 Cas de mots isolés

Nous avons testé notre phonétiseur dans le cas des mots isolés avec les corpus de test présentés ci-dessus : le dictionnaire *MCAW-Dict* (Tableau 4.3), et le corpus des noms propres (Tableau 4.4) et les résultats étaient très satisfaisants. Le score PER était de 3.1% et le WER de 12.5%.

La mesure PER est calculée par l'expression suivante :

$$PER = \frac{\text{Nombre de phonèmes erronés}}{\text{Nombre total de phonèmes}} \times 100$$

Pour la norme WER, est calculée comme suit :

$$WER = \frac{\text{Nombre de mots phonétisés par TOP-AS contenant au moins une erreur de transcription}}{\text{Nombre total de mots}} \times 100$$

4.3.2.2 Cas de textes

Pour prendre en compte le contexte phonético-phonologique, nous avons également testé notre phonétiseur sur un corpus de plusieurs textes soigneusement sélectionnés et phonétisés manuellement, comme mentionné à la section 4.3.1. Ce sont quelques chapitres du Saint Coran, plus 1400 phrases épuisées de différentes sources textuelles, pour un total de 12000 mots. Le Tableau 4.5 donne un aperçu de ce corpus phonétique de test. Nous avons présenté également en Tableau 4.6 un autre extrait de ce corpus avec des passages de poésie pré-islamique.

Les résultats testés sont mesurés en termes de PER et de taux d'erreur en termes de phrases (SER : Sentence Error Rate). Ces résultats sont présentés dans le Tableau 4.7. *Bi-gramme* signifie que nous prenons une séquence de deux graphèmes de part et d'autre de la transition entre les mots. De même, *3-gramme* signifie que l'on prend une séquence de trois graphèmes de part et d'autre de la transition entre les mots, etc., et ceci pour inclure l'information sur les phénomènes phonético-phonologiques qui interviennent au niveau des transitions entre les mots.

La mesure SER est calculée comme suit :

$$SER = \frac{\text{Nombre de phrases phonétisées par TOP-AS avec au moins un phonème incorrect}}{\text{Nombre total de phrases}} \times 100$$

Tableau 4.7. Score de la TOP sur un corpus textuel phonétisé

	Bi-Gram	3-gram	4-gram
PER (%)	15.4	10.7	5.3
SER (%)	40.2	21.8	14.5

4.3.3 Analyse des Erreurs

Il existe différentes dimensions selon lesquelles nous pouvons souhaiter classer les erreurs de transcription. Une première information importante est d'obtenir une liste de jeux de mots qui ont entraîné une erreur de transcription. Un deuxième type d'information est lié à la catégorie de l'origine de l'erreur, ou finalement à la nature même du phonème erroné. Notre système TOP-AS comprend plusieurs sous-modules, qui effectuent des analyses linguistiques et phonétiques spécifiques, telles que la segmentation, le prétraitement, la syllabation, etc. Évidemment, identifier quel sous-module est responsable de quelle erreur était une de nos préoccupations. En ce qui concerne spécifiquement le prétraitement, c'est un fait que les caractéristiques typographiques d'un jeu de mot (comporte-t-il des chiffres arabes ou romains ? est-il un mot composé, ...) détermine souvent son traitement dans le système. De plus, la classification des erreurs selon cette dimension spécifique fournit également des indications utiles sur la robustesse du système face à des entrées typographiques inattendues, et est par conséquent très pertinente dans le cadre d'une analyse détaillée. A notre connaissance, très peu de tentatives ont été faites pour définir une grille raisonnable de classification et d'analyse des erreurs de transcription en AS. Dans la *figure 4.1* nous exposons les principales catégories des erreurs décelées de notre système phonétiseur TOP-AS. Ce graphe est le résultat d'une étude minutieuse des cas les plus problématiques de défaillance du système. La *figure 4.2* donne les taux d'erreurs de transcription par phonème.

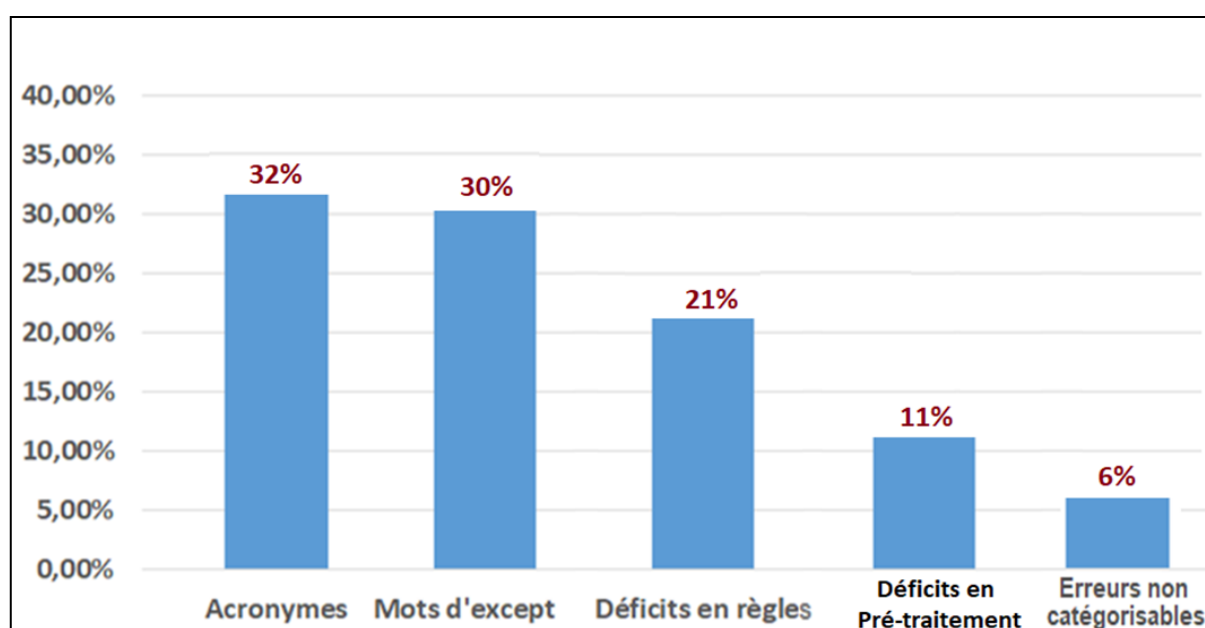


Fig. 4.1 : Catégories d'erreurs décelées du système TOP-AS

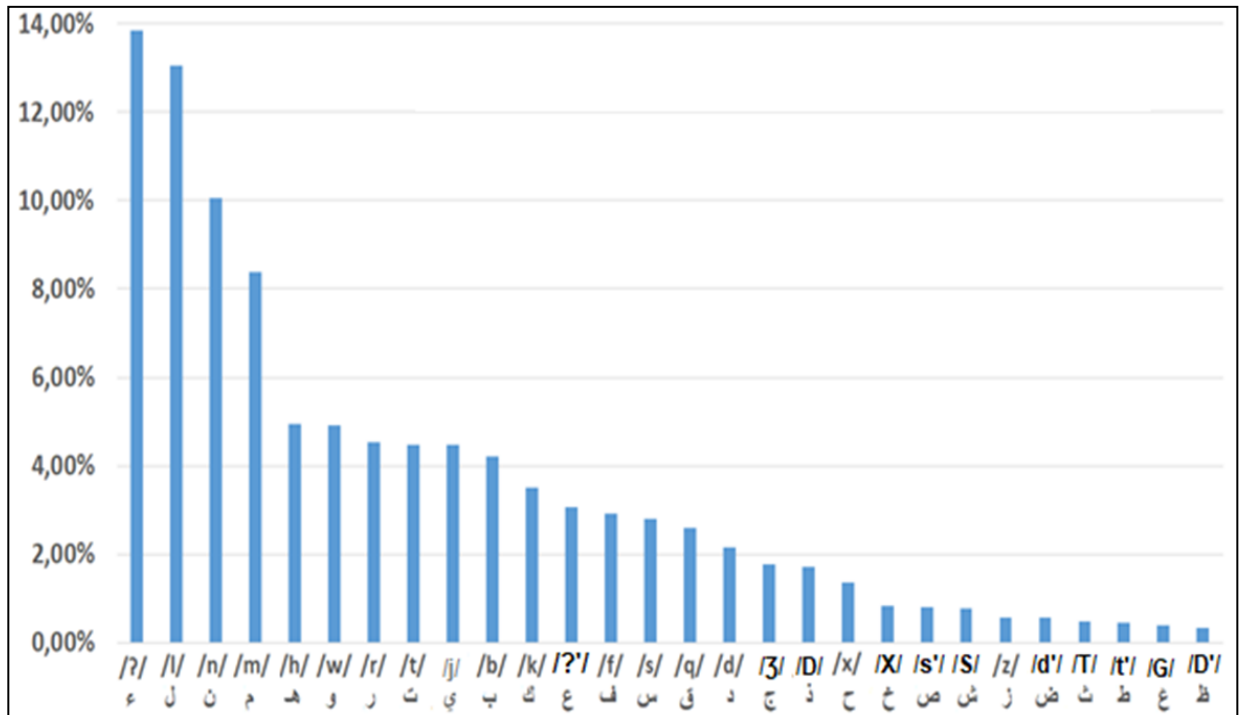


Fig. 4.2 : Taux d'erreurs de transcription par phonème

4.4 Conclusion

Nous avons présenté dans ce chapitre les différentes techniques utilisées pour l'évaluation des phonétiseurs pour l'Anglais, le Français et pour l'AS. Nous avons exposé ensuite les ressources de test que nous avons construit pour le besoin d'évaluation de notre système de transcription. Enfin, nous avons exposé l'approche adoptée pour l'évaluation de notre système de transcription. Les résultats obtenus étaient tout à fait satisfaisants attestant l'efficacité de notre approche. Le meilleur score obtenu par notre phonétiseur dans le cas de mots isolés, en termes de WER est de 12,5 %. Cela veut dire que 12,5% de mots phonétisés par notre système comportent au moins une erreur de transcription. Dans le cas de textes, le meilleur score obtenu par notre phonétiseur, en termes de SER est de 14,5 %. Cela veut dire que 14,5% de phrases phonétisées par notre système comportent au moins une erreur de transcription.

Ces erreurs se partagent entre des phénomènes connus (mots d'exception ou acronymes non répertoriés dans les bases de données, ...) et une multitude de problèmes locaux difficilement catégorisables. Ces derniers problèmes, certainement dus au manque de couverture de nos bases de règles, ne peuvent être traités que par une analyse minutieuse des corpus traités.

Conclusion Générale et Perspectives

La transcription *Orthographique-Phonétique* constitue un composant nécessaire et une étape primordiale pour l'élaboration d'un système de synthèse de la parole à partir du texte. Après avoir formulé les problèmes posés par la langue arabe tant au niveau orthographique qu'au niveau phonétique et phonologique, nous avons élaboré des règles exhaustives de transcription qui couvrent tous les phénomènes analysés ainsi que les exceptions. Avec de tels éléments d'analyse, nous avons proposé un système à base de règles qui a été vérifié à l'aide de corpus bien choisis. Notre système de Transcription Orthographique-Phonétique est un automate paramétré appliquant une large gamme de règles de réécriture, qui permettent d'associer un phonème (ou un groupe de phonèmes) à un caractère (ou un groupe de caractères) orthographique, en prenant en compte le contexte gauche (caractères ou groupes de caractères précédant le segment à transcrire) et le contexte droit (caractères ou groupes de caractères suivant le segment à transcrire). Ces règles sont organisées de façon hiérarchique, des règles les plus particulières aux règles les plus générales.

La construction d'un système de phonétisation automatique à base de règles nécessite une bonne connaissance de la langue et de ses règles de phonétisation qui, par ailleurs, ne doivent pas contenir trop d'exceptions : la phonétisation automatique utilisant des connaissances linguistiques ne nécessite pas de techniques innovantes, mais plutôt une expertise importante sur la langue. L'avantage de cette approche est qu'elle permet de mieux contrôler la qualité de la construction des dictionnaires de prononciation : en cas d'erreur il est possible d'ajouter une nouvelle règle. De plus, le coût de développement peut être comparé au coût nécessaire à la construction manuelle du corpus d'apprentissage utilisé par les approches statistiques (Data-Driven : Approches guidées par les données).

Les résultats des tests – tant pour les mots isolés que pour le texte- ont montré que la précision de notre système de transcription est très élevée. La précision de la conversion lettre-son est fonction du prétraitement, de la précision des règles « lettre-son » et de l'exhaustivité du lexique des mots à orthographe irrégulière, des abréviations, des symboles et des acronymes irréguliers.

Une discussion approfondie sur la phonétisation du texte arabe basée sur des règles a été présentée et plusieurs extensions et améliorations de l'état de l'art ont été proposées. Une contribution majeure de cette étude est l'ensemble de règles proposé qui traite des phénomènes phonétiques et phonologiques de l'AS tant à l'intérieur du mot, qu'entre les bords des mots dans le cas de la parole continu, tels que la pharyngalisation, emphatisation, gémination, nasalisation, et d'autres phénomènes de coarticulation anticipatoire. Certains de ces phénomènes -comme la pharyngalisation- est subtils et difficiles à prendre en compte formellement.

Notre système phonétiseur pour l'AS, baptisé *TOP-AS*, exploite une large panoplie de ressources numériques comme la Base de Données des abréviations, le lexique des mots d'exception, le lexique des acronymes, la liste des mots-outils les plus répandus en AS. *TOP-AS* a été testé sur un large dictionnaire phonétique de plus de 35000 mots de vocabulaire avec leurs phonétisations, en plus d'un corpus important de noms propres, et sur un corpus de plusieurs textes soigneusement sélectionnés et phonétisés manuellement, formé de quelques chapitres du Saint Coran, plus 1400 phrases épuisées de différentes sources textuelles, pour un total de 12000 mots. Nous nous sommes efforcés de choisir des textes contenant un assortiment complet des difficultés typiques de la conversion G2P : des structures compliquées, citations, dates, noms propres, acronymes et fautes de frappe. De plus, une transcription phonémique de ce type de corpus de textes a été unanimement jugée d'une grande pertinence pour la communauté de chercheurs dans le domaine de la synthèse de la parole.

Les résultats montrent que l'approche proposée donne une précision supérieure à 97% en termes d'erreurs en phonèmes (PER), une précision supérieure à 88% en termes d'erreurs en mots (WER), et une précision supérieure à 85% en termes d'erreurs en phrases (SER). Le score de précision en termes de phrases (SER) est plus significatif que celui en termes de mots (WER). En effet, il était beaucoup plus facile pour nous de trouver un accord sur une définition exploitable d'une phrase que d'un mot (pensez aux élisions, aux abréviations, aux nombres et aux mots composés).

Concernant les erreurs recensées après tests et évaluations des performances de notre système de phonétisation *TOP-AS*, nous pouvons y remédier en généralisant les cas les plus fréquents par des règles manquantes, ou des méthodes d'apprentissage, et en codant les autres cas dans les listes d'exceptions, telles sont les tâches qu'il nous faut continuer à mener pour améliorer le système.

Comme travail futur, nous projetons :

- tout d'abord d'améliorer les performances du processus TOP en intégrant un module d'analyse morpho-syntaxique, et ceci en collaboration avec la filière « Traitement Automatique de la Langue Arabe » du département d'informatique ;
- intégrer, dans la phase de prétraitement, un processus de repérage et d'identification des expressions numériques, et leurs conversions en mots en AS, et ceci par une approche à base d'algorithmes génétiques et d'apprentissage automatique ;
- ensuite, développer un moteur TTS en AS qui s'appuie sur les résultats présentés dans cette étude, et ceci après avoir tester l'intégration de notre phonétiseur dans les systèmes TTS existants -prenants en charge la langue AS- tels que eSpeak, Festival et MBROLA.

Les travaux futurs immédiats incluront la conception et la réalisation d'une plate-forme crowdsourcing (application collaborative) qui sera hébergée sous GitHub pour l'aide à la translittération et à la transcription phonétique incluant tous les types de codification (*IPA*, *SAMPA*, *Buckwalter*) afin de produire des ressources numériques en AS nécessaires au Traitement Automatique de la Parole.

Références Bibliographiques

- [1] Cole R. A. et al. Survey of the State of the Art in Human Language Technology, Cambridge University Press and Giardini, 1997. (Web Edition, <http://www.dfki.de/~hansu/HLTSurvey.pdf>).
- [2] Flanagan, J. Rabiner L. Speech Synthesis. Dowden, Hutchinson & Ross, Inc., Pennsylvania, 1973.
- [3] Allen S., Hunnicut S., Klatt D. From Text To Speech, The MITTALK System. Cambridge University Press, 1987.
- [4] Dutoit, T. An Introduction to Text-to-Speech Synthesis, KLUWER Academic Publishers, Dordrecht, 1997.
- [5] Kleijn K., Paliwal K. Speech Coding and Synthesis. Elsevier Science B.V., The Netherlands, 1998.
- [6] Santen J., Sproat R., Olive J., Hirschberg J. Progress in Speech Synthesis, Springer Verlag New York Inc. (Includes CDROM), 1997.
- [7] Grainger, J., Ferrand, L. Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of memory and language*, 35(5), 1996, pp. 623-647.
- [8] Black, Alan W., Kevin, A. Lenzo. Building Synthetic Voices, Language Technologies Institute, Carnegie Mellon University, 2003, (<http://festvox.org/bsv/>).
- [9] Flanagan, J. Speech Analysis, Synthesis, and Perception. Springer Verlag, Berlin Heidelberg New York, 1972.
- [10] Donovan, R. Trainable Speech Synthesis. PhD. Thesis. Cambridge University, Engineering Department, England. 1996, http://svr-ftp.eng.cam.ac.uk/pub/reports/donovan_thesis.ps.Z
- [11] O'Saughnessy, D. Speech Communication Human and Machine. Addison Wesley, 1987.
- [12] Witten, I. Principles of Computer Speech, Academic Press Inc, 1982.
- [13] Veldhuis R., Bogaert I., Lous N. Two-Mass Models for Speech Synthesis. *Proceedings of Euro-speech 95* (3), 1995, pp. 1853-1856.
- [14] Klatt, D. Review of Text-to-Speech Conversion for English. *Journal of the Acoustical Society of America*, JASA vol. 82 (3), 1987, pp. 737-793.
- [15] Schroeder M. A Brief History of Synthetic Speech. *Speech Communication*, vol. 13, 1993, pp. 231-237.
- [16] Guth G., Burgtähler D. Synthèse De La Parole, *Compte Rendu D'une Séance De Travail Sur L'électrical Vocal Analog (eva III)*. Esquisse D'une Méthode. *AL-Lisaniyyat*, Volume 1, Numéro 1, 1971, pp. 7-62.
- [17] Barber S. et al. A Rule Based Italian Text-to-Speech System. *Proceedings of Eurospeech 89*, 1989 (1), pp. 517-520.
- [18] Carlson, R. Evaluation and Development of the KTH Text-to-Speech System on the Segmental Level. *Proceedings of ICASSP 90* (1), 1990, pp. 317-320.
- [19] Karlsson, I., Neovius, L. Speech Synthesis Experiments with the GLOVE Synthesizer. *Proceedings of Eurospeech 93* (2), 1993, pp. 925-928.
- [20] Lee, K. Hidden Markov Models: Past, Present, and Future. *Proceedings of Euro-speech 89* (1): 1989, pp. 148-155.
- [21] Leipp, E., Castellengo, M., Liénard, J.S. La synthèse de la parole à partir de digrammes phonétiques, 6th Int. Cong. on Acoust., Tokyo, août 1968.
- [22] Fant, G. Glottal flow: Models and interaction. *Journal of Phonetics* 14(3-4), 1986, pp. 393-399. DOI: 10.1016/S0095-4470(19)30714-4
- [23] Hadjipantelis, P. Z. Functional Data Analysis in Phonetics, Phd Thesis, Department of Statistics, University of Warwick, 2013.
- [24] Haton, J. P., Pierrel, J.M., Perennou, G., Caelen, J., et Gauvain, J.L. Reconnaissance automatique de la parole. Dunod Informatique, Paris, 1991.

- [25] Al-Kheir, J., Smekal, Z. Cepstral Vocal Tract Modelling for Text-To-Speech Synthesis. *Damascus University Journal*, Vol. (29) - No. (1) 2013, pp. 11-20.
- [26] Markel, J.D., Gray, A.H. Jr. *Linear Prediction of Speech*. Springer Science & Business Media, 12 mars 2013, 290 pages.
- [27] Fant, G. *Acoustic theory of speech production*, The Hague, The Netherlands, Mouton, 1960.
- [28] LIBAL, U., JOHANSSON, K. H. Yule-Walker equations using higher order statistics for nonlinear autoregressive model. In : *2019 Signal Processing Symposium (SPSymposium)*. IEEE, 2019. pp. 227-231.
- [29] El Jaroudi, A., Makhoul, J. M. Discrete all-pole modelling. *IEEE Transactions on Signal Processing*, volume 39, 1991, pp. 411–423.
- [30] Edwards, T. J. Multiple features analysis of intervocalic English plosives. *Journal of Acoustical Society of America*, volume 69, 1981, pp. 535–547.
- [31] Di Cristo, A. La prosodie au carrefour de la phonétique, de la phonologie et de l'articulation formes fonctions. *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix en Provence (TIPA), Laboratoire Parole et Langage*, 2004, 23, pp. 67-211. hal00285554.
- [32] Chomsky, N., Halle, M. *The sound pattern of English*, New York, Harper & Row, 1968.
- [33] Jakobson R., Fant C.G. M., Halle, M. *Preliminaries to Speech Analysis*. Technical Report No. 13 Acoustics Laboratory, M.I.T 1952.
- [34] Calliope, P., Fant, G. *La parole et son traitement automatique*, Masson, 1989.
- [35] Le Meur, P.Y. *Synthèse de parole par unités de taille variable*. Ph. D. thesis, Ecole Nationale Supérieure des Télécommunications (ENST), 1996.
- [36] Naïli, A., Chebbine, B., Boudraa, M., Caelen G. H. Proposition d'un outil pour l'étude de la prosodie de la langue arabe. *Colloque Maghrébin d'Automatique et d'Electronique, Tlemcen (Algérie)*, Vol. 2, 1996, pp 131 – 136.
- [37] Baloul, S. *Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé*, Thèse de Doctorat, Université du Maine, 2003.
- [38] Klatt, D. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.*, 67(3), pp. 971–995, March 1980.
- [39] Tsukanova, A. *Articulatory speech synthesis*. Thèse de doctorat, Ecole doctorale IAEM Lorraine, 2019.
- [40] Prudon, R., Alessandro, C. A selection/concatenation TTS synthesis system: Databases development, system design, comparative evaluation. *4th Speech Synthesis Workshop, Pitlochry, Scotland*, 2001, pp. 137-142.
- [41] Dutoit, T., et al. Synthèse de la parole et reconnaissance de la parole: droites gauches et monde parallèles, *Actes du 6^{ème} Congrès Français d'Acoustique*, 2002, pp. 8-11.
- [42] Charpentier, F., Moulines, E. TTS Algorithms Based on FFT Synthesis. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* 88, 1988, pp. 667–670.
- [43] Alessandro, C. et al. Synthèse par sélection : prosodie, dialogue et qualité vocale. *Cahiers de l'Institut de Linguistique de Louvain*, Peeters, 2004, 30, pp.153-180. (hal02009065)
- [44] Sproat, R. et al. Multilingual text analysis. Chapitre 3, In book: *Multilingual Text-To-Speech Synthesis: The Bell Labs Approach*, Publisher: Kluwer, 1997, pp. 31–87.
- [45] Chodorowski, J. *Grammatical inference for learning syntax in speech recognition and oral dialogue*, Thesis, University of Rennes I, 2001, P50.
- [46] Boite, R. et al. *Traitement de la parole*, Presses polytechniques et universitaires romandes, 2000.
- [47] Herment, S. Relations entre prosodie et syntaxe. *Anglo-ponia/Sigma*, En ligne depuis 20 Avril 2015, accès le 11 Novembre 2020. URL: <http://journals.openedition.org/angloponia/402> ; DOI : <https://doi.org/10.4000/angloponia.402>
- [48] Botinis, A. *Intonation: Analysis, Modelling and Technology*. Springer, 2000. DOI: 10.1007/9789401143172.

- [49] Fujisaki, H. Modelling the Process of Fundamental Frequency Contour Generation. *Speech Perception, Production and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis Bateson, Y. Sagisaka, IOS Press, 1992, pp. 313-328.
- [50] Bartkova, K. and C. Sorin. A model of segmental duration for speech synthesis in French. *Speech Communication*, 1987, 6, pp. 245–260.
- [51] Goubanova, O., Taylor, P. Using Bayesian Belief networks for model duration in text-to-speech systems. In *CDROM Proc. ICSLP 2000*, Beijing, China, 2000.
- [52] Stöber, K., Portele, T., Wagner, P., Hess, W. Synthesis by word concatenation. *Proceeding of Sixth European Conference on Speech Communication and Technology, EUROSPEECH'99*, Budapest, Hungary, September 5-9, 1999, pp. 619-622.
- [53] TABET, Y., BOUGHAZI, M. Speech synthesis techniques. A survey. In : *International Workshop on Systems, Signal Processing and their Applications, WOSSPA*. IEEE, 2011. pp. 67-70.
- [54] Bigorgne, D. et al. Multilingual PSOLA Text-to-Speech System. *Proceedings of the IEEE Conf. on Acoustics Speech and Signal Processing II*, 1993, pp. 187–190.
- [55] Klabbers, E., Veldhuis, R. Reducing Audible Spectral Discontinuities. *IEEE Transactions on Speech and Audio Processing* 9 (1), 2001, pp. 39-51, DOI: 10.1109/89.890070
- [56] Boeffard, O., Cherbonnel, Emerard, F., and White, S. Automatic segmentation and quality evaluation of speech unit inventories for concatenation-based multilingual PSOLA test to speech systems. *Euro-speech '93*, 1993, pp. 1449–1452.
- [57] Hunt, A. J., Black, A. W. Unit selection in a concatenative speech synthesis system using a large speech database. *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, Atlanta, GA, USA, vol. 1, 1996, pp. 373-376. doi: 10.1109/ICASSP.1996.541110.
- [58] Sagisaka, Y. Speech synthesis by rule using an optimal selection of non-uniform synthesis units. *IEEE Int. Conf. Acoust., Speech, Signal Proc.*, New York, 1988, pp. 679-682.
- [59] Syrdal, A.K., Wightman, C.W., Conkie, A., Stylianou, Y., Beutnagel, M., Schroeter, J., Strom, V., Lee, K.S., and Makashay, M.J. Corpus-based techniques in the AT&T NextGen synthesis system. In: *Proceedings of the International Conference on Spoken Language Processing (Beijing)*, 3, 2000, pp. 410–413.
- [60] Syrdal, A., Möhler, G., Dusterhoff, K., Conkie, A., and Black, A. Three methods of intonation modeling. *Proceedings of the Third International Workshop on Speech Synthesis (Jenolan Caves, Australia)*, 1998, pp. 305–310.
- [61] Santen J., Buchsbaum A. Methods for optimal text selection. *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech'97)*, 1997, pp. 553-556, Rhodes, Greece.
- [62] Boëffard, O., Miclet, S. & White, S. Automatic generation of optimized unit dictionaries for Text-to-Speech synthesis. *Proceedings on the International Conference on Speech and Language Processing'92*, Banff, 1992, pp. 1211– 1214
- [63] Taylor, P. A., Isard, S. D. Automatic diphone segmentation. *Proceedings on Eurospeech'91*, Genoa, 1991, pp. 709–711.
- [64] Balestri, M. et all. Choose the best to modify the least: a new generation concatenative synthesis system. *EURO-SPEECH'99*, 1999, pp. 2291-2294.
- [65] Campbell, N. CHATR: A High-Definition Speech Re-sequencing System. In *Proceedings of ASA/ASJ Joint Meeting* (pp. 1223-1228). December 23-28, 1996, Hawaii, USA.
- [66] Dutoit, T., Leich, H. MBRPSOLA: Text-to-Speech synthesis based on an MBE re-synthesis of the segments database. *Speech Communication* 13, (34), 1993, pp. 435-440.

- [67] Stylianou, Y. Applying the harmonic plus noise model in concatenative speech synthesis, in *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 1, 2001, pp. 21-29. doi: 10.1109/89.890068.
- [68] Beijer, L. Evaluating the Suitability of Orthographic Transcription and Intelligibility Scale Rating of Semantically Unpredictable Sentences (SUS) for Speech Training Efficacy Research in Dysarthric Speakers with Parkinson's Disease. *Journal of Medical Speech Language Pathology*, **20**, 2012, pp. 17-34.
- [69] Demostenes, Z. et al. Intelligent Learning Techniques applied to Quality Level in Voice over IP Communications. *International Journal on Advances in Internet Technology*, vol 6 no 3 & 4, 2013, pp 145—155. http://www.iariajournals.org/internet_technology/
- [70] Letowski, Tomasz R., Angelique, A. Scharine. Correlational Analysis of Speech Intelligibility Tests and Metrics for Speech Transmission, US Army Research Laboratory, ARLTR8227, Approved for public release DEC 2017.
- [71] Howard, P. J. SOAP, Speech Output Assessment Package, Version 4.0, ESPRIT SAMUCL042, 1992.
- [72] Sindran, F., Mualla, F., Bobzin, K., Nöth, E. Automatic robust rule-based phonetization of standard Arabic. *Text, Speech, and Dialogue*, Vol. 9302 of LNAI, Springer, 2015, pp. 442–451.
- [73] Hunnicutt, S., Meng, H., Seneff, S., Zue, V. Reversible letter-to-sound sound-to-letter generation based on parsing word morphology. *Proceedings of 3rd European Conference on Speech Communication and Technology (Eurospeech '93)*, Vol.2, Berlin, Germany, 1993, pp. 763–766.
- [74] Boula de Mareuil P. et al. Evaluation of Grapheme-to-Phoneme Conversion for Text-to-Speech Synthesis in French. 1st International Conference on Language Resources and Evaluation, 1998 Granada.
- [75] Yvon, F. Objective evaluation of grapheme-to-phoneme conversion for text-to-speech synthesis in French. *Computer Speech and Language*, 12(4), 1998, pp. 393-410.
- [76] Hassan, Z. M., Esling, John H. Laryngoscopic articulatory and acoustic evidence of a prevailing emphatic feature over the word in arabic, *ICPhS XVI*, ID 1280, Saarbrücken, 610 August 2007, pp. 1753-1756.
- [77] Cherifi, E-H., Guerti, M. Arabic grapheme-to-phoneme conversion based on joint multi-gram model. *Int J Speech Technol.* 2021, vol. 24, n°1, pp. 173–182. ISSN 1381-2416. <https://doi.org/10.1007/s10772-020-09779-8>
- [78] Cherifi, E-H., Guerti, M. Conditional Random Fields Applied to Arabic Orthographic-Phonetic Transcription. *Archives of Acoustics*. Vol. 46, No. 2, 2021, pp. 237-247. <https://doi.org/10.24425/aoa-2021.136574>
- [79] Coker, C. H. A dictionary-intensive letter-to-sound program, *J. Acoust. Soc. Amer.* 78 Suppl. 1, S7, 1985.
- [80] Coker, C. H., Church, K. W., Liberman, M. Y. Morphology and Rhyming: Two powerful alternatives to Letter-to-Sound Rules for speech Synthesis. *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans (France), 1990, pp. 83-86.
- [81] Levinson, S.E., Olive, J.P., Tschirgi, J.S. *Speech Synthesis in Telecommunications*. *IEEE Communications Magazine*, November 1993, pp. 46-53.
- [82] Laporte, E. Méthodes algorithmiques et lexicales de phonétisation de textes. Applications au Français. Thèse de doctorat, Paris, Université Paris 7, 1988.
- [83] AlFairuz Abadi, M. M. Al Qamous Almuhiy. Arresalah Publishers, 1996, Beirut.
- [84] Ainsworth, W.A. A system for converting English text into speech. *IEEE Transactions on Audio and Electro-acoustics*, 1973, AU21, pp. 288–290.
- [85] McIlroy, M. *Synthetic English Speech by Rule*, Bell Telephone Laboratories Memo, 1974.
- [86] A.M. Harrison, W.Y. Lau, H. Meng and L. Wang. Improving Mispronunciation Detection and Diagnosis of Learners' Speech with Context-Sensitive Phonological Rules based on Language Transfer. In *Proceedings of InterSpeech 2008*, Brisbane, Australia, pp. 2787 – 2790

- [87] Hertz, S.R. Appropriateness of different rule types in speech synthesis. In: Wolf, J.J., Klatt, D.H. (Eds.), *Speech Communication Papers*, No. 50, Acoustical Society of America, 1979, pp. 511–514.
- [88] Hunnicutt, S. Grapheme to phoneme rules: a review. *Speech Transmission Laboratory Quarterly Progress and Status Report*, Royal Institute of Technology (KTH), Stockholm, STL- QPSR 2–3/1980, pp. 38–60.
- [89] Belrhali, R.V. et al. From lexicon to rules: towards a descriptive of french text-to-phonetics transcription. *Proceedings of the International Conference on Spoken Language Processing 92*, 1992, Alberta, pp. 1183–1186.
- [90] Divay, M., Vitale, A.J. Algorithms for grapheme–phoneme translation for English and French: applications for database searches and speech synthesis. *Computational Linguistics* 23 (4), 1997, pp. 495–523.
- [91] Beesley, K. Arabic Morphological Analysis on the Internet. *Proceedings of the Sixth International Conference and Exhibition on Multilingual Computing*, 1998, Cambridge, England.
- [92] Shaalan, K. Rule-based Approach in Arabic Natural Language Processing. *Int. Jour. on Inf. and Com.*, Vol. 3, No. 3, June 2010, pp. 11-19.
- [93] Elnaggar, A. A finite state automata of the Arabic grammar. *Journal of Natural Language Processing*, 1993, 19 (9).
- [94] Hassanain, A.T., Shahata, H. *Rules for Arabic writing: Theory and Applications*, Arabic Book House, Le Caire (en arabe), 1998.
- [95] Qazzi, N.M. *Summary of Arabic writing rules*, Qarib Publications Establishment, Le Caire (en arabe), 2000.
- [96] Damper, R.I., Marchand, Y., Adamson, M.J., Gustafson, K. Evaluating the pronunciation component of Text-to-Speech systems for English: a performance comparison of different approaches. *Computer Speech and Language* 13 (2), 1999, pp. 155–176.
- [97] De Mareuil, P., Yvon, F., D’Alessandro, C., Auberge, V., et al. Evaluation of grapheme-to-phoneme conversion for Text-to-Speech synthesis in French. In: *Proceedings of the First International Conference on Language Resources and Evaluation*, Grenade, 1998, pp. 641–646.
- [98] Van den Bosch, A., Daelemans, W. Data-oriented methods for grapheme-to-phoneme conversion, *EACL '93: Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics*, 1993, pp. 45–53. <https://doi.org/10.3115/976744.976751>
- [99] Dedina, M., Nusbaum, H. PRONOUNCE : a program for pronunciation by analogy. *Computer Speech & Language*. 5(1), 1991, pp. 55-64. 10.1016/08852308(91)90017K.
- [100] Yvon, F. *Prononcer par analogie : motivation, formalisation et évaluation = pronunciation by analogy: motivations, implementations and evaluations*, PhD thesis, Ecole nationale supérieure des télécommunications, Paris, France, 1996.
- [101] Yvon, F. Paradigmatic cascades: a linguistically sound model of pronunciation by analogy. In: *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Spain, 1997, pp. 429–435.
- [102] Damper, R.I., Eastmond, J.F.G. Pronunciation by analogy: impact of implementational choices on performance. *Language and Speech* 40 (1), 1997, pp. 1–23.
- [103] Bagshaw, P.C. Phonemic transcription by analogy in Text-to-Speech synthesis: novel word pronunciation and lexicon compression. *Computer Speech and Language*, 1998, 12 (2), pp. 119–142.
- [104] Marchand, Y., Damper, R.I. A multi-strategy approach to improving pronunciation by analogy. *Computational Linguistics* 26 (2), 2000, pp. 195–219.
- [105] Damper, R.I. *Data-Driven Methods in Speech Synthesis*. Kluwer Academic Publishers, Dordrecht, 2001.
- [106] Luk, R., Damper, R. Stochastic phonographic transduction for English. *Computer Speech and Language* 10, 1996, pp. 133–153.

- [107] Daelemans, W., van den Bosch, A., Weijters, T. IGTre: using trees for compression and classification in lazy learning algorithms. *Artificial Intelligence Review* 11, 1997, pp. 407–423.
<https://doi.org/10.1023/A:1006506017891>
- [108] Sejnowski, T., Rosenberg, C.R. Parallel networks that learn to pronounce English text. *Complex Systems* 1, 1987, pp. 145–168.
- [109] Matsumoto, T., Yamaguchi, Y. A multi-language text-to-speech system using neural networks. In: Bailly, G., Benoit, C. (Eds.), *Proceedings of the ESCA Workshop on Speech Synthesis*. ESCA, Atrants, 1990, pp. 269–272.
- [110] Cherifi E. H., Guerti M. Phonetisaurus-based letter-to-sound transcription for Standard Arabic. 5th International Conference on Electrical Engineering-Boumerdes (ICEE-B), 2017, Boumerdes. Publisher: IEEE, pp. 1-4. Doi : 10.1109/ICEE-B.2017.8192073
<https://ieeexplore.ieee.org/abstract/document/8192073>
- [111] Meng, H.M. Phonological parsing for bidirectional letter-to-sound and sound-to-letter generation, Ph.D. thesis, MIT Press, Cambridge, MA, 1995.
- [112] Balabaki, R. Arabic and Semitic Writing Systems: Studies into the Origins and History of Semitic Writing [en arabe]. The Science for Millions, 1981, Beirut.
- [113] Sampson, G. *Writing Systems*. Hutchinson, London, 1985.
- [114] Gibbon, D., Moore, R., Winski, R. *Handbook of Standard and Resources for Spoken Language Systems*. Mouton de Gruyter, Berlin, 1997.
- [115] ALL, Ikbel Hadj, MNASRI, Z., LAACHRI, Z. Gemination prediction using DNN for Arabic text-to-speech synthesis. In : *2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)*. IEEE, 2019. pp. 366-370.
- [116] Haralambous, Y. Tour du monde des ligatures. *Cahiers Gutenberg, Association GUTenberg, Ligatures & caractères contextuels*, 22, 1995, pp.85-97.
- [117] Saidane T. et all. La transcription orthographique phonétique de la langue arabe. *Proceedings of RECITAL. Fès. 2004*.
- [118] Schmidt, M., Fitt, S., Scott, C., Jack, M. Phonetic transcription standards for European names (ONOMASTICA). *Proceedings of the Eurospeech*, vol. 93. ESCA, Berlin, 1993, pp. 279–282.
- [119] Belhoula, K. Rule-based grapheme-to-phoneme conversion of names. *Proceedings of the Euro-speech*, 1993, pp. 881–884.
- [120] Vitale, T. An algorithm for high accuracy name pronunciation by parametric speech synthesizer. *Computational Linguistics* 17, 1991, pp. 257–276.
- [121] Bartlett S. et al. On the Syllabification of Phonemes, *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, Boulder, Colorado, pp. 308–316.
- [122] ABDO, M. S., KANDIL, A. H. Semi-automatic segmentation system for syllables extraction from continuous Arabic audio signal. *International Journal of Advanced Computer Science and Applications*, 2016, vol. 7, no 1, pp. 535-540.
- [123] Anis, I. *Language Phonetics (en Arabe)*. The Anglo-Egyptian Book Publishers, 1992, Le Caire.
- [124] Abbes, R. La conception et la réalisation de concordancier électronique pour l’arabe. Thèse de Doctorat, Institut national des sciences appliquées de Lyon, Lyon, France, 2004.
- [125] McAllister, M. The problem of punctuation ambiguity in full automatic text-to-speech conversion. In: *Proceedings of the Euro-speech 89*, vol. 1, Paris, 1989, pp. 538–541
- [126] Liberman, M.Y., Church, K.W. Text analysis and word pronunciation in text-to-speech systems. In: Furui, S., Sohndi, M. (Eds.), *Advances in Speech Signal Processing*. Dekker, New York, 1992, pp. 791–831.
- [127] Refenstette, G., Tapanainen, P. What is a word, what is a sentence? Problems of tokenization. *The 3rd International Conference on Computational Lexicography*, Budapest, 1994, pp. 79–87.

- [128] Doumi, N. Extraction de connaissances à partir du texte, thèse de doctorat, Université Djillali Liabes, Faculté des Sciences Exactes, Sidi Bel Abbes, 2017.
- [129] Van Leeuwen, H.C. Speech maker formalism: a rule formalism operating on a multi-level, synchronized data structure. *Computer Speech and Language*, vol. 7(4), 1993, pp. 369-390.
- [130] Dutoit, T., et al. Euler: an open, generic, multilingual and multiplatform Text-to-Speech system. In: *Proceedings of LRECO00*, Athens, 2000, pp. 563–566
- [131] Humoud, M. *The Encyclopedia of Arabic Spelling and Grammar*. Dar Al fikr Al libnani, Beirut, 1998.
- [132] Humoud, M. *The Encyclopedia of Arabic Names and Their Meanings*. Dar Al fikr Al libnani, Beirut, 1995.
- [133] Bin-Muqbil, Musaed S. *Phonetic and phonological aspects of arabic emphatics and gutturals*, Phd Thesis in Linguistics, university of wisconsin-madison, 2006.
- [134] Embarki M., Guillemot Ch., Barkatdefradas M. Expansion nasale en arabe standard : indices acoustiques d'une coarticulation anticipatoire, *Revue Parole*, Vol. 3940, 2006, pp. 209-234.
- [135] Sawalha, M., Brierley, C., Atwell, E. Automatically generated, phonemic Arabic IPA pronunciation tiers for the Boundary Annotated Qur'an Dataset for Machine Learning. *Proceeding of 2nd Workshop on Language Resources and Evaluation for Religious Texts*, Iceland, 2014, pp. 42-47.
- [136] Rajouani, A. et al. Synthesis of the pharyngealization feature in Arabic. *Speech Communication* 6 (3), North Holland, 1987, pp. 261-268.
- [137] ElImam, Y.A. Speech synthesis using partial syllables. *Computer Speech and Language* 4, 1990, pp. 203–229.
- [138] Shane, S. *Generative phonology*. Englewood Cliffs, New Jersey. Prentice Hall, INC, 1973.
- [139] Ladefoged, P. *A course in phonetics*. U.S.A: Hein and Heinle, 1993.
- [140] Fromkin, V. and Rodman, R. *An Introduction to Language*. (4th ed), U.S.A: Holt, Rinehart and Winston, Inc, 1988.
- [141] Al Jamal, Abdurrahman Yousuf. *Mukhtasar Al Mughni fi Ulm Attajweed*. Ghaza : The Islamic University, 2000.
- [142] Nassir, A. Q. *Ghayat Al Mureed fi Ulm Attajweed*. Jeddah, Kunooz Al Ma'rifa Library, 1997.
- [143] Issam M. Abu Salim. Consonant assimilation in Arabic: An auto-segmental perspective, *Lingua*, Volume 74, Issue 1, January 1988, pp. 45-66.
- [144] Alduais, Ahmed Mohammed S. *Quranic Phonology and Generative Phonology: Formulating Generative Phonological Rules to Non-Syllabic Nuun's Rules*. *International Journal Of Linguistics*, 2013, Vol. 5, No. 5, pp. 33-61.
- [145] Roach, P. *English Phonetics and Phonology*. 3rd ed. Longman: Cambridge UP, 1987.
- [146] RAHIM, A. J., *et al.* Assimilation in Classical Arabic. *Adab AL Rafidayn*, 2013, vol. 43, no 66.
- [147] Morel, M. et al. Utilisation d'une structure arborescente pour une hiérarchisation fine des règles de transcription graphème-phonème, congrès AFCP, 1997, pp.151-154.
- [148] Bechet, F. *LIA_phon, un système complet de phonétisation de textes*. 2001.
- [149] Jekosch, U., *Voice and Speech Quality Perception: Assessment and Evaluation Signals and Communication Technology*, 2005.
- [150] Bernstein, J., Nessly, L. Performance comparison of component algorithms for the phonemization of orthography. *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, CA, 1981, pp. 19–21.
- [151] Kucera, H., Francis, W.N. *Computational Analysis of Present-Day American English*. Brown University Press, Providence, RI, 1967.

- [152] Novak, J.R., Minematsu, N., Hirose, K. WFST-based grapheme-to-phoneme conversion: open source tools for alignment, model-building and decoding. 10th International Workshop on Finite State Methods and Natural Language Processing, 2012, p. 45-49.
- [153] Galescu, L., Allen, J.F. Pronunciation of proper names with a joint n-gram model for bi-directional grapheme-to-phoneme conversion. 7th International Conference on Spoken Language Processing, 2002, pp. 109–112.
- [154] Kheang, S., Katsurada, K., Iribe, Y., Nitta, T. Solving the phoneme conflict in grapheme-to-phoneme conversion using a two-stage neural network-based approach. IEICE Trans. Inf. Syst. 97(4), 2014, pp. 901–910.
- [155] Wu, K., Allauzen, C., Hall, K., Riley, M., Roark, B. Encoding linear models as weighted finite-state transducers. Interspeech 2014, ISCA, pp. 1258-1262.
- [156] Rao, K., Peng, F., Sak, H., Beaufays, F. Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2015, pp. 4225-4229.
- [157] Brigitte B. SPPAS : un outil « userfriendly » pour l’alignement texte/son. JEPTALNRECITAL 2012, Atelier DEGELS 2012 : Défi GEstE Langue des Signes, pp. 85–92, Grenoble, 4 au 8 juin 2012.
- [158] Alessandro, C. et al. Evaluation de la synthèse de parole à partir du texte dans la francophonie : premiers résultats, Ressources et évaluation en Ingénierie des langues. Duculot, Bruxelles, 2000.
- [159] Selim, H., Anbar, T. A phonetic transcription system of Arabic text. Acoustics, Speech, and Signal Processing. IEEE International Conference on ICASSP'87. IEEE, 1987, pp. 1446-1449.
- [160] Ellmam, Y. A. An unrestricted vocabulary Arabic speech synthesis system. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 37(12), 1989, pp. 1829–1845.
- [161] Ahmed, M.E. Toward an Arabic text-to-speech system. The Arabian Journal for Science and Engineering, 1991, 16(4), 565-583.
- [162] Al-ghamdi, M., Al-muhtasib, H. and Elshafei, M. Phonetic Rules in Arabic Script. Journal of King Saud University, Computer and Information Sciences, 2004, 16, pp. 85-115.
- [163] Biadisy, F. et al. Improving the Arabic pronunciation dictionary for phone and word recognition with linguistically-based pronunciation rules. Proceed. of Hum. Lang. Tech: The Annual Conference of the North American Chapter of the Asso. for Comput. Linguistics, 2009, pp. 397–405.
- [164] Habash, N., Rambow, O., Roth, R. Mada+tokan: a toolkit for Arabic tokenization, diacritization, morphological disambiguation, pos tagging, stemming and lemmatization. In: Proceedings of the 2nd International Conference on Arabic Language Resources and Tools (MEDAR), Cairo, Egypt, 2009, pp. 102–109.
- [165] Al-Ghamdi, M., Elshafei, M., Al-Muhtaseb, H. Arabic broadcast news transcription system. Int. J. Speech Technol. 10 (4), 2014, pp. 183–195.
- [166] Ramsay A., Alsharhan I., Ahmed H. Generation of a Phonetic Transcription for Modern Standard Arabic: A Knowledge-Based Model. Computer Speech & Language, 28(4), 2014, pp. 959-978.
- [167] Cherifi E. H. (2020). MCAW-Dict, Phonetic Dictionary of the *Most Commonly used Arabic Words* with SIMPA Transcription, document électronique disponible sur: https://drive.google.com/file/d/1h_dPwAXKone7nGIKgelMt8mlzGYFF7d2/view?usp=sharing
- [168] Atwell, E. et all. Understanding the Quran: A New Grand Challenge for Computer Science and Artificial Intelligence. ACM/BCS Visions of Computer Science, 2010.

Annexe 1

Classification des sons de l'AS selon leur point d'articulation

Alphabet/Son	Prononciation (transcrite en IPA)	point d'articulation
ء	/ʔalif/	Laryngale occlusive
ب	/ba:ʔ/	Labiale occlusive sonore
ت	/ta:ʔ/	Dentale occlusive sourde
ث	/θ:ʔ/	Interdentale émise en insérant le bout de la langue entre les dents ; spirante sourde
ج	/dʒi:m/	Spirante palatale sonore
ح	/ħa:ʔ/	Spirante laryngale sourde
خ	/ħa:ʔ/	Vélaire spirante sourde
د	/da:l/	Dentale occlusive sonore
ذ	/ða:l/	Interdentale spirante sonore émise en insérant le bout de la langue entre les dents.
ر	/ra:ʔ/	Vibrante linguale souvent appelée « liquide »
ز	/za:j/	Dentale spirante sonore
س	/si:n/	Dentale spirante sourde
ش	/ʃi:n/	Palatale spirante sourde
ص	/ʃa:d/	Emphatique; dentale spirante sourde vélarisée
ض	/ɢa:d/	Emphatique; inter dentale spirante sonore vélarisée
ط	/ɢa:ʔ/	Emphatique ; dentale occlusive sourde vélarisée
ظ	/ðʕa:ʔ/	Emphatique; inter dentale spirante sonore vélarisée
ع	/ʕajn/	Laryngale spirante sonore
غ	/ɣajn/	Vélaire spirante sonore
ف	/fa:ʔ/	Labiodentale spirante sourde
ق	/qa:f/	Occlusive arrière-vélaire sourde accompagnée d'une explosion glottale
ك	/ka:f/	Palatale occlusive sourde
ل	/la:m/	Linguale souvent appelée « liquide »
م	/mi:m/	Labiale nasalisée
ن	/nu:n/	Dentale nasalisée
ه	/ħa:ʔ/	Spirante sonore
و	/wa:w/	Consonne /Voyelle longue vélaire labiale
ي	/ja:ʔ/	Consonne /Voyelle longue palato-alvéolaire

Annexe 2 : SAMPA

1- SAMPA (Direct), version étendue

ء	/ʔ/
ب	/b/
ت	/t/
ث	/T/
ج	/ʒ/, /dʒ/, /g/
ح	/x/
خ	/X/
د	/d/
ذ	/D/, /ð/
ر	/r/
ز	/z/
س	/s/
ش	/S/, /ʃ/
ص	/sʰ/

ض	/dʰ/
ط	/tʰ/
ظ	/Dʰ/
ع	/ʔʰ/, /ʔn/
غ	/G/, /ɣ/
ف	/f/
ق	/q/
ك	/k/
ل	/l/
م	/m/
ن	/n/
هـ	/h/
و	/w/
ي	/j/

Fatha فتحة	/a/
Voyelle longue : Alif ألف ممدودة	/a:/
Kasra كسرة	/i/
Voyelle longue : Ya ياء ممدودة	/i:/
Damma ضمة	/u/
Voyelle longue : Waw واو ممدودة	/u:/

[tanwi:n fatxa] تنوين فتحة	/an/
[tanwi:n d'amma] تنوين ضمة	/un/
[tanwi:n kasra] تنوين كسرة	/in/

2- SAMPA avec exemples

Symbol SAMPA	Exemple transcrit	Notation	Orthographe En arabe
Consonnes			
Les plosives			
b	ba:b	door	بَابُ
t	tis?`	nine	تسع
d	da:r	home	دار
t`	t`a:bi?`	stamp	طابع
d`	d`arab	he hit	ضرب
k	kabi:r	large	كبير
g	gami:l (en prononciation Egyptienne) dʒami:l Ou zami:l en prononciation standard	beautiful	جميل
ʒ ou dʒ	ʒabalun ou dzabalun	Mountain	جبل
ʔ	ʔakl	food	أكل
q	qalb	heart	قلب
Fricatives			
f	fi:l	elephant	فيل
v	nivi:n	Nevien (personal name)	نفين
T	Tala:T	three	ثلاث
D	Dakar	male	ذكر
D`	D`ala:m	darkness	ظلام
s	sa?`i:d	happy	سعيد
z	zami:l	colleague	زميل
s`	s`aGi:r	small	صغير
S	Sams	sun	شمس
Z	Zami:l	beautiful	جميل
X	Xit`a:b	letter	خطاب
G	Garb	west	غرب
X\	X\ilm	dream	حلم
?` (?\)	?`alam	flag	علم
h	hawa:ʔ	air	هواء
Nasales			
m	ma:l	money	مال
n	nu:r	light	نور

Trille			
r	rima:l	sand	رمال
Latérales			
l	la:	no	لا
lʔ	?alʔah	God	الله
Semi-voyelles			
w	wa:hid	one	واحد
j	jawm	day	يوم
Voyelles			
i	Dʔil	shadow	ظل
a	Xal	solution	حل
u	?ʔumr	age	عمر
i:	?ʔi:d	feast	عيد
a:	ma:l	money	مال
u:	fu:l	beans	فول

Annexe 3

Tableau étendue des graphèmes arabes avec leurs codes :
ASCII, UNICODE et translittération BUCKWALTER

Unicode Dec.	Unicode Hex.	Graph.	Code BUCKWALTER	Orthographe
1569	U+0621	ء	'	Hamza
1571	U+0623	أ	>	Alif + HamzaAbove
1572	U+0624	ؤ	&	Waw + HamzaAbove
1573	U+0625	إ	<	Alif + HamzaBelow
1574	U+0626	ئ	}	Ya + HamzaAbove
1575	U+0627	ا	A	Alif
1576	U+0628	ب	b	Ba
1577	U+0629	ة	p	TaMarbuta
1578	U+062A	ت	t	Ta
1579	U+062B	ث	v	Tha
1580	U+062C	ج	j	Jeem
1581	U+062D	ح	H	HHa
1582	U+062E	خ	x	Kha
1583	U+062F	د	d	Dal
1584	U+0630	ذ	*	Thal
1585	U+0631	ر	r	Ra
1586	U+0632	ز	z	Zain
1587	U+0633	س	s	Seen
1588	U+0634	ش	\$	Sheen
1589	U+0635	ص	S	Sad
1590	U+0636	ض	D	DDad
1591	U+0637	ط	T	TTa
1592	U+0638	ظ	Z	DTha

1593	U+0639	ع	E	Ain
1594	U+063A	غ	g	Ghain
1600	U+0640	-	-	Tatweel
1601	U+0641	ف	f	Fa
1602	U+0642	ق	q	Qaf
1603	U+0643	ك	k	Kaf
1604	U+0644	ل	l	Lam
1605	U+0645	م	m	Meem
1606	U+0646	ن	n	Noon
1607	U+0647	ه	h	Ha
1608	U+0648	و	w	Waw
1609	U+0649	ى	Y	AlifMaksura
1610	U+064A	ي	y	Ya
1611	U+064B	؀	F	Fathatan
1612	U+064C	؁	N	Dammatan
1613	U+064D	؂	K	Kasratan
1614	U+064E	؃	a	Fatha
1615	U+064F	؄	u	Damma
1616	U+0650	؅	i	Kasra
1617	U+0651	؆	~	Shadda
1618	U+0652	؇	o	Sukun
1619	U+0653	؈	^	Maddah
1620	U+0654	؉	#	HamzaAbove
1648	U+0670	ا	`	AlifKhanjareeya
1649	U+0671	آ	{	Alif + HamzatWasl
1756	U+06DC	س	:	SmallHighSeen

1759	U+06DF	◦	@	SmallHighRoundedZero
1760	U+06E0	◦	"	SmallHighUprightRectangularZero
1762	U+06E2	◡	[SmallHighMeemIsolatedForm
1763	U+06E3	س	;	SmallLowSeen
1765	U+06E5	و	,	SmallWaw
1766	U+06E6	ا	.	SmallYa
1768	U+06E8	ن	!	SmallHighNoon
1770	U+06EA	◦	-	EmptyCentreLowStop
1771	U+06EB	◦	+	EmptyCentreHighStop
1772	U+06EC	◦	%	RoundedHighStopWithFilledCentre
1773	U+06ED	◡]	SmallLowMeem

Annexe 4 : Traitement des nombres

1- Les chiffres romains de 1 à 1000

1	I	34	XXXIV	90	XC	199	CXCIX
2	II	35	XXXV	98	XCVIII	200	CC
3	III	36	XXXVI	99	XCIX	300	CCC
4	IV	37	XXXVII	100	C	301	CCCI
5	V	38	XXXVIII	101	CI	400	CD
6	VI	39	XXXIX	109	CIX	401	CDI
7	VII	40	XL	110	CX	500	D
8	VIII	41	XLI	111	CXI	501	DI
9	IX	42	XLII	119	CXIX	600	DC
10	X	43	XLIII	120	CXX	700	DCC
11	XI	44	XLIV	121	CXXI	702	DCCII
12	XII	45	XLV	129	CXXIX	703	DCCIII
13	XIII	46	XLVI	130	CXXX	723	DCCXXIII
14	XIV	47	XLVII	131	CXXXI	750	DCCL
15	XV	48	XLVIII	139	CXXXIX	751	DCCLI
16	XVI	49	XLIX	140	CXL	770	DCCLXX
17	XVII	50	L	141	CXLI	771	DCCLXXI
18	XVIII	51	LI	148	CXLVIII	800	DCCC
19	XIX	59	LIX	149	CXLIX	825	DCCCXXV
20	XX	60	LX	150	CL	850	DCCCL
21	XXI	61	LXI	151	CLI	851	DCCCLI
22	XXII	62	LXII	159	CLIX	860	DCCCLX
23	XXIII	67	LXVII	160	CLX	861	DCCCLXI
24	XXIV	69	LXIX	161	CLXI	900	CM
25	XXV	70	LXX	166	CLXVI	949	CMXLIX
26	XXVI	71	LXXI	170	CLXX	950	CML
27	XXVII	78	LXXVIII	171	CLXXI	951	CMLI
28	XXVIII	79	LXXIX	179	CLXXIX	970	CMLXX
29	XXIX	80	LXXX	180	CLXXX	971	CMLXXI
30	XXX	81	LXXXI	181	CLXXXI	982	CMLXXXII
31	XXXI	82	LXXXII	190	CXC	983	CMLXXXIII
32	XXXII	83	LXXXIII	197	CXCVII	984	CMLXXXIV
33	XXXIII	89	LXXXIX	198	CXCVIII	1000	M

2- Code en python de conversion des chiffres romains en chiffres arabes :

```
# -*- coding: utf-8 -*-
import codecs
import sys
import re

numerals = [
    {'letter': 'M', 'value': 1000},
    {'letter': 'D', 'value': 500},
    {'letter': 'C', 'value': 100},
    {'letter': 'L', 'value': 50},
    {'letter': 'X', 'value': 10},
    {'letter': 'V', 'value': 5},
    {'letter': 'I', 'value': 1},
]

def roman_to_arabic(number):
    index_by_letter = {}
    for index in range(len(numerals)):
        index_by_letter[numerals[index]['letter']] = index

    result = 0
    previous_value = None
    for letter in reversed(number):
        index = index_by_letter[letter]
        value = numerals[index]['value']
        if (previous_value is None) or (previous_value <= value):
            result += value
        else:
            result -= value
        previous_value = value

    return result

main()
```

3- Code en python de conversion des chiffres en mots arabes

L'objectif est de fournir une fonction à usage général qui soit simple mais précise pour convertir les expressions numériques en mots arabes.

Exemples:

```
Num2Word_AR(2000) // "ألفان"  
Num2Word_AR(15000120) // "خمسة عشر مليونًا ومائة وعشرون"  
Num2Word_AR(2020) // "ألفان وعشرون"  
Num2Word_AR(2020) // "ألفان وعشرون"  
Num2Word_AR(100) // "مئة"  
Num2Word_AR(200) // "مئتان"  
Num2Word_AR(350) // "ثلاثمئة وخمسون"
```

La sortie :

```
ألفان  
خمسة عشر مليونًا ومائة وعشرون  
ألفان وعشرون  
مئة  
مئتان  
ثلاثمئة وخمسون
```

```

# -*- coding: utf-8 -*-

import re
from decimal import Decimal
from math import floor

CURRENCY_DZD = [ ("دينارا", "ديناران", "دنانير", "دينارًا"),
                 ("سنتيم", "اثنين سنتيم", "سنتيمات", "سنتيم") ]

ARABIC_ONES = [
    "واحد", "اثنان", "ثلاثة", "أربعة", "خمسة", "ستة", "سبعة", "ثمانية",
    "تسعة",
    "عشرة", "أحد عشر", "اثنان عشر", "ثلاثة عشر", "أربعة عشر", "خمسة عشر",
    "ستة عشر", "سبعة عشر", "ثمانية عشر",
    "تسعة عشر"
]

class Num2Word_AR(object):
    errmsg_too_big = "Too large"
    max_num = 10 ** 36

    def __init__(self):
        self.number = 0
        self.arabicPrefixText = ""
        self.arabicSuffixText = ""
        self.integer_value = 0
        self._decimalValue = ""
        self.partPrecision = 2
        self.currency_unit = CURRENCY_DZD[0]
        self.currency_subunit = CURRENCY_DZD[1]
        self.isCurrencyPartNameFeminine = True
        self.isCurrencyNameFeminine = False
        self.separator = 'و'

        self.arabicOnes = ARABIC_ONES
        self.arabicFeminineOnes = [
            "إحدى", "اثنان", "ثلاث", "أربع", "خمس", "ست", "سبع", "ثمان",
            "تسع",
            "عشر", "إحدى عشرة", "اثنان عشرة", "ثلاث عشرة", "أربع عشرة",
            "خمس عشرة", "ست عشرة", "سبع عشرة", "ثمانية عشرة",
            "تسع عشرة"
        ]

```

```

self.arabicOrdinal = [
    "اول", "ثاني", "ثالث", "رابع", "خامس", "سادس", "سابع", " ", " ",
    "ثامن",
    "تاسع", "عاشر", "حادي عشر", "ثاني عشر", "ثالث عشر", "رابع
    عشر",
    "خامس عشر", "سادس عشر", "سابع عشر", "ثامن عشر", "تاسع عشر"
]
self.arabicTens = [
    "عشرون", "ثلاثون", "أربعون", "خمسون", "ستون", "سبعون", "
    ثمانون",
    "تسعون"
]
self.arabicHundreds = [
    "مائة", "مئتان", "ثلاثمائة", "أربعمائة", "خمسمائة", " ", " ",
    "ستمائة",
    "سبعمائة", "ثمانمائة", "تسعمائة"
]
self.arabicAppendedTwos = [
    "مئتا", "ألفا", "مليون", "مليار", "تريليون", "
    كوادريليون",
    "كوينتليون", "سكستليون"
]
self.arabicTwos = [
    "مئتان", "ألفان", "مليونان", "ملياران", "تريليونان",
    "كوادريليونان", "كوينتليونان", "سكستليونان"
]
self.arabicGroup = [
    "مائة", "ألف", "مليون", "مليار", "تريليون", "كوادريليون",
    "كوينتليون", "سكستليون"
]
self.arabicAppendedGroup = [
    "ألفاً", "مليوناً", "ملياراً", "تريليوناً", "كوادريليوناً",
    "كوينتليوناً", "سكستليوناً"
]
self.arabicPluralGroups = [
    "آلاف", "ملايين", "مليارات", "تريليونات", " ", " ",
    "كوادريليونات",
    "كوينتليونات", "سكستليونات"
]

def number_to_arabic(self, arabic_prefix_text, arabic_suffix_text):
    self.arabicPrefixText = arabic_prefix_text
    self.arabicSuffixText = arabic_suffix_text
    self.extract_integer_and_decimal_parts()

```

```

def extract_integer_and_decimal_parts(self):
    re.split('\.\.', str(self.number))
    splits = re.split('\.\.', str(self.number))

    self.integer_value = int(splits[0])
    if len(splits) > 1:
        self._decimalValue = int(self.decimal_value(splits[1]))
    else:
        self._decimalValue = 0

def decimal_value(self, decimal_part):

    if self.partPrecision is not len(decimal_part):
        decimal_part_length = len(decimal_part)

        decimal_part_builder = decimal_part
        for i in range(0, self.partPrecision -
decimal_part_length):
            decimal_part_builder += '0'
            decimal_part = decimal_part_builder

        if len(decimal_part) <= self.partPrecision:
            dec = len(decimal_part)
        else:
            dec = self.partPrecision
            result = decimal_part[0: dec]
    else:
        result = decimal_part

    for i in range(len(result), self.partPrecision):
        result += '0'
    return result

def digit_feminine_status(self, digit, group_level):
    if group_level == -1:
        if self.isCurrencyPartNameFeminine:
            return self.arabicFeminineOnes[int(digit)]
        else:
            return self.arabicOnes[int(digit)]
    elif group_level == 0:
        if self.isCurrencyNameFeminine:
            return self.arabicFeminineOnes[int(digit)]
        else:
            return self.arabicOnes[int(digit)]
    else:
        return self.arabicOnes[int(digit)]

```

```

def process_arabic_group(self, group_number, group_level,
                        remaining_number):
    tens = Decimal(group_number) % Decimal(100)
    hundreds = Decimal(group_number) / Decimal(100)
    ret_val = ""

    if int(hundreds) > 0:
        if tens == 0 and int(hundreds) == 2:
            ret_val = "{}".format(self.arabicAppendedTwos[0])
        else:
            ret_val =
"{}".format(self.arabicHundreds[int(hundreds)])

    if tens > 0:
        if tens < 20:
            if tens == 2 and int(hundreds) == 0 and group_level > 0:
                if self.integer_value in [2000, 2000000, 2000000000,
20000000000000, 20000000000000000, 20000000000000000000]:
                    ret_val = "{}".format(
                        self.arabicAppendedTwos[int(group_level)])
                else:
                    ret_val = "{}".format(
                        self.arabicTwos[int(group_level)])
            else:
                if ret_val != "":
                    ret_val += " و "

                if tens == 1 and group_level > 0 and hundreds == 0:
                    ret_val += ""
                elif (tens == 1 or tens == 2) and (
                    group_level == 0 or group_level == -1) and \
                    hundreds == 0 and remaining_number == 0:
                    ret_val += ""
                else:
                    ret_val += self.digit_feminine_status(int(tens),
group_level)
            else:
                ones = tens % 10
                tens = (tens / 10) - 2
                if ones > 0:
                    if ret_val != "" and tens < 4:
                        ret_val += " و "

                    ret_val += self.digit_feminine_status(ones,
group_level)
                if ret_val != "" and ones != 0:
                    ret_val += " و "

                ret_val += self.arabicTens[int(tens)]

    return ret_val

```

```

def convert(self, value):
    self.number = "{:.9f}".format(value)
    self.number_to_arabic(self.arabicPrefixText,
self.arabicSuffixText)
    return self.convert_to_arabic()

def convert_to_arabic(self):
    temp_number = Decimal(self.number)

    if temp_number == Decimal(0):
        return "صفر"

    decimal_string = self.process_arabic_group(self._decimalValue,
                                                -1,
                                                Decimal(0))

    ret_val = ""
    group = 0

    while temp_number > Decimal(0):

        number_to_process = int(
            Decimal(str(temp_number)) % Decimal(str(1000)))
        temp_number = int(Decimal(temp_number) / Decimal(1000))

        group_description = \
            self.process_arabic_group(number_to_process,
                                      group,
                                      Decimal(floor(temp_number)))

        if group_description != '':
            if group > 0:
                if ret_val != "":
                    ret_val = "{} و {}".format("", ret_val)
                if number_to_process != 2:
                    if number_to_process % 100 != 1:
                        if 3 <= number_to_process <= 10:
                            ret_val = "{} {}".format(
                                self.arabicPluralGroups[group], ret_val)
                        else:
                            if ret_val != "":
                                ret_val = "{} {}".format(
                                    self.arabicAppendedGroup[group],
                                    ret_val)
                            else:
                                ret_val = "{} {}".format(
                                    self.arabicGroup[group],
                                    ret_val)
            else:
                ret_val="{} {}".format(self.arabicGroup[group],
ret_val)
                ret_val = "{} {}".format(group_description, ret_val)
                group += 1

```



```

formatted_number = ""
    if self.arabicPrefixText != "":
        formatted_number += "{} ".format(self.arabicPrefixText)
    formatted_number += ret_val
    if self.integer_value != 0:
        remaining100 = int(self.integer_value % 100)

        if remaining100 == 0:
            formatted_number += self.currency_unit[0]
        elif remaining100 == 1:
            formatted_number += self.currency_unit[0]
        elif remaining100 == 2:
            if self.integer_value == 2:
                formatted_number += self.currency_unit[1]
            else:
                formatted_number += self.currency_unit[0]
        elif 3 <= remaining100 <= 10:
            formatted_number += self.currency_unit[2]
        elif 11 <= remaining100 <= 99:
            formatted_number += self.currency_unit[3]
    if self._decimalValue != 0:
        formatted_number += " {}".format(self.separator)
        formatted_number += decimal_string

    if self._decimalValue != 0:
        formatted_number += " "
        remaining100 = int(self._decimalValue % 100)

        if remaining100 == 0:
            formatted_number += self.currency_subunit[0]
        elif remaining100 == 1:
            formatted_number += self.currency_subunit[0]
        elif remaining100 == 2:
            formatted_number += self.currency_subunit[1]
        elif 3 <= remaining100 <= 10:
            formatted_number += self.currency_subunit[2]
        elif 11 <= remaining100 <= 99:
            formatted_number += self.currency_subunit[3]

    if self.arabicSuffixText != "":
        formatted_number += " {}".format(self.arabicSuffixText)

    return formatted_number

def validate_number(self, number):
    if number >= self.max_num:
        raise OverflowError(self.errmsg_too_big)
    return number

def set_currency_prefer(self, currency):
    if currency == 'DZD':
        self.currency_unit = CURRENCY_DZD[0]
        self.currency_subunit = CURRENCY_DZD[1]

```

```

def to_currency(self, value, currency='DZD', prefix='', suffix=''):
    self.set_currency_prefer(currency)
    self.isCurrencyNameFeminine = False
    self.separator = "و"
    self.arabicOnes = ARABIC_ONES
    self.arabicPrefixText = prefix
    self.arabicSuffixText = suffix
    return self.convert(value=value)

def to_ordinal(self, number, prefix=''):
    if number <= 19:
        return "{}".format(self.arabicOrdinal[number])
    if number < 100:
        self.isCurrencyNameFeminine = True
    else:
        self.isCurrencyNameFeminine = False
    self.currency_subunit = ('', '', '', '')
    self.currency_unit = ('', '', '', '')
    self.arabicPrefixText = prefix
    self.arabicSuffixText = ""
    return "{}".format(self.convert(abs(number)).strip())

def to_year(self, value):
    value = self.validate_number(value)
    return self.to_cardinal(value)

def to_ordinal_num(self, value):
    return self.to_ordinal(value).strip()

def to_cardinal(self, number):
    number = self.validate_number(number)
    minus = ''
    if number < 0:
        minus = 'سالب '
    self.separator = ','
    self.currency_subunit = ('', '', '', '')
    self.currency_unit = ('', '', '', '')
    self.arabicPrefixText = ""
    self.arabicSuffixText = ""
    self.arabicOnes = ARABIC_ONES
    return minus + self.convert(value=abs(number)).strip()

```