

SYNTHAR+ :
Un Système de Pré-traitement de Textes Arabes
en Vue de leur Synthèse Orale sous le Système Multivox

Z. Zemirli^{*el**}, N. Vigouroux^{**}, M. Sellami^{***}

^{*}*I.N.I Institut National d'Informatique* BP 68 M Oued-smar 16270 Alger Algérie
(email : cipap@ist.cerist.dz - zemirli@irit.fr)

^{**}*I.R.I.T., IHMPT, UPS* - 118, rte de Narbonne F-31062 Toulouse cedex France
(email : vigourou@irit.fr)

^{***}*Université d'Annaba - Institut d'Informatique* - Annaba RP 23000 Algérie
(email : sellami@ist.cerist.dz)

Résumé:

Cet article présente le module SYNTHAR+ de pré-traitements linguistiques, destiné à une bonne oration des textes de la langue Arabe sous le système de synthèse Multivox. Il repose sur la modélisation de connaissances phonologiques et morpho-lexicales pour élaborer la chaîne de graphèmes. Un protocole d'évaluation illustrant l'amélioration de l'intelligibilité de la restitution est enfin discuté.

Mots clés : synthèse de textes, connaissances phonologiques et morpho-lexicales, langue arabe, système Multivox, mesures de l'intelligibilité.

Abstract :

This article presents the SYNTHAR+ linguistic pre-processing tool which aim is to secure the phonetisation performance of the Arabic texts under the Multivox System. It is based on the phonological and morpho-lexical knowledge modelling to produce the graphemes string. An evaluation protocol describing the increasing of the text-to-speech intelligibility is then discussed.

key words : Text-to-speech synthesis, phonological and morpho-lexical knowledge, arabic language, Multivox system, intelligibility measures.

Introduction

Les systèmes de synthèse de textes trouvent de plus en plus leur utilisation dans des applications de serveurs vocaux, de systèmes de lectures de documents électroniques, etc. Des erreurs de phonétisation ont été identifiées et partiellement résolues par le développement de pré-traitements linguistiques sur la chaîne à phonétiser pour la langue française [Cotto 92], anglaise [McAlister 89], allemande [Belhoula 93], [Traber 93], suédoise [Backstrom 89]. Concernant la langue arabe, à ce jour, il n'existe pas de système commercialisé qui prenne en entrée, une chaîne orthographique en langue arabe pour générer une restitution sonore bien que beaucoup de travaux de laboratoires aient été menés [Guerti 85], [El Shaffei 88], [Najim 91], [Ghazali 92a] et [Ghazali 92b].

Le système de synthèse multilingue Multivox de la société hongroise AKADIMPEX, étendu aux spécificités de la langue arabe comprend une base de 40 sons et nécessite pour fonctionner en entrée la chaîne phonétique exacte à restituer.

Dans une première partie, cet article fait état de la nécessité de prendre en compte des connaissances phonologiques dans les processus de phonétisation de textes en langue arabe. Le point suivant concerne l'apport des connaissances phonologiques et morpho-lexicales en tant qu'outil d'aide à la vocalisation automatique de textes arabes. Ces deux sources de connaissance sont intégrées dans le système SYNTHAR+.

Nous rappelons brièvement les principes de base des systèmes de synthèse vocale ainsi que les caractéristiques et les limitations de Multivox. Le traitement des inattendus (sigles, abréviations, noms propres, etc.) et de quelques problèmes phonologiques tels que la gémination des consonnes et la liaison d'articles sont également discutés. Les principes et les connaissances mises en œuvre dans SYNTHAR+ sont décrits. Enfin une évaluation de l'intelligibilité de la synthèse sous Multivox avec la contribution ou non des connaissances phonologiques et morpho-lexicales est analysée.

1. Les systèmes de synthèse vocale

Il existe deux techniques de synthèse de la parole: la synthèse vocale à partir de parole codée et la synthèse à partir de texte. Ces deux méthodes sont très différentes tant au niveau des traitements qu'elles mettent en œuvre que des applications qui en découlent. Pour plus d'informations, le lecteur pourra se reporter à [Charpentier 89], [Sorin 91], [Rigoni 96].

1.1. Synthèse numérique codée

Ce mode de synthèse est basé sur l'enregistrement numérique de séquences de mots ou de phrases prononcées par un locuteur. Elles sont stockées et assemblées au moment de la restitution de façon à produire la suite d'éléments sonores de la phrase à synthétiser. Cette technique a l'avantage de produire une restitution vocale quasi-naturelle avec une bonne prosodie. Par contre, elle possède un inconvénient majeur car elle nécessite une capacité de stockage très importante. Des techniques de compression associées aux CD-ROM permettent de réduire considérablement cet encombrement. La restitution de la parole codée ne peut être utilisée que dans des applications où la taille du vocabulaire est limitée et, où les messages vocaux sont connus a priori.

1.2. Synthèse à partir de texte

Cette technique s'appuie sur une représentation orthographique du texte à synthétiser. A partir d'une suite de caractères écrits, la carte de synthèse génère la suite de sons correspondants. La restitution vocale est dans ce cas, beaucoup moins naturelle et intelligible qu'avec la technique de synthèse à partir de la parole codée. Le principal avantage reste néanmoins que l'on peut, grâce à ce mode de synthèse restituer n'importe quel texte écrit [Sorin 91], [Cotto 92]. La figure 1 illustre les principaux constituants d'un système de synthèse à partir de textes, à savoir :

- un module de transcription graphème-phonème d'une chaîne orthographique en une chaîne phonétique incluant généralement un module de pré-traitements linguistiques plus ou moins élaboré
- et un module de génération acoustique et prosodique.

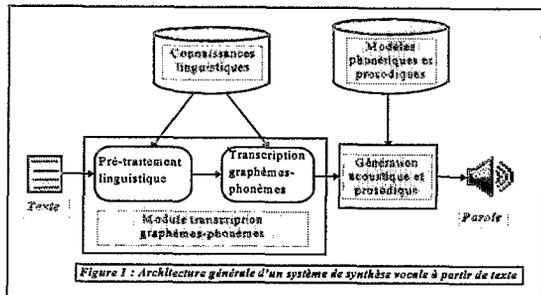


Figure 1 : Architecture générale d'un système de synthèse vocale à partir de texte

La transcription d'un texte orthographique en un texte phonétique comporte des difficultés dues à la non correspondance entre les symboles orthographiques et les symboles phonétiques. Les calculs des contours prosodiques ont quant à eux besoin d'informations sur la structure syntaxique voir même

sémantique d'une phrase pour être correctement effectués. On se reportera à [Aggoun 85], [Rossi 87], [Rossi 88], [Rossi 89] et à [El Shaffei 88] pour des études faites sur les langues française et arabe. Dans cet article, nous nous intéressons principalement aux procédés de transcription de la chaîne orthographique voyellisée en la chaîne phonétique. Auparavant, nous allons décrire brièvement le système de synthèse Multivox pour la langue arabe.

2. Caractéristiques du système de synthèse vocale Multivox

Le système de synthèse Multivox [Multivox 94] permet de transcrire une chaîne phonétique latine (correspondant au texte arabe) en une succession de phonèmes arabes à synthétiser. Le tableau 1 établit la correspondance des 40 graphèmes arabes et des phonèmes connus du système Multivox. Cette bibliothèque de phonèmes comprend les 26 consonnes, les 3 voyelles brèves (a) َ, (o) ُ et (i) ِ ainsi que les géminées de certaines consonnes [Zemirli 94a]. Les voyelles longues (ee) ا (oo) و ou (ii) ي n'ont pas de correspondant dans le système Multivox. Elles font l'objet d'un traitement morpho-orthographique (cf. § 3.2.1.1).

1	2	1	2	1	2	1	2	1	2	1	2
s"	ص	e'	ء أ	k	ك	d'	ض	d	د	e	ا
d"	ض	e	ـ	l	ل	t'	ط	z'	ذ	b	ب
t"	ظ	i	ـ	m	م	z"	ظ	r	ر	t	ت
zz"	ظ	o	ـ	n	ن	o:	ع	z	ز	c	ث
o"	غ	h"	ح	h	ه	g	غ	sz	س	j	ج
		zz'	ذ	w	و	f	ف	s	ش	h'	ح
		ssz	س	y	ي	q	ق	s'	ص	x	خ

Tableau 1 : Correspondance Graphèmes Arabes (2) - Phonèmes Multivox (1)

Le tableau 2 illustre des écritures de mots dans l'alphabet du système Multivox.

bileedon	بلاد	e:esszemeee'	السماء	fes'lon	فصلن
e'elxesebo	الخشبة	mismis	مشمش	yerbit'o	يربط

Tableau 2 : Exemple d'écriture dans « l'alphabet » Multivox

Les premiers essais d'utilisation du système Multivox [Zemirli 94a] nous amènent à constater les contraintes d'utilisation suivantes :

- « l'alphabet » en entrée du système Multivox est composée de caractères latins,
- les mots abrégés et les sigles ne sont pas correctement synthétisés,
- les dates et les chiffres sont complètement ignorés,
- les ponctuations autres que : « !, . et ? » sont ignorées,
- le système commence la synthèse dès la rencontre d'un point, d'un point d'interrogation ou un point d'exclamation. Les longues phrases sont synthétisées sans interruption provoquant ainsi une perte du sens qu'elle véhicule. Un module métatextuel est nécessaire pour découper le texte en phrase à synthétiser.

Si la chaîne à synthétiser est : *يَكْتُبُ عَمْرٌ رِسَالَةً إِلَى صَدِيقِهِ مُصَنَّفِي.* (Omar a écrit une lettre à son ami Mustapha.) alors la chaîne nécessaire à l'entrée du synthétiseur Multivox devra être : *ketebe o:omero riszeeleten e'ilee s'ediqiqihi mos't'efee.*

3. Description de SYNTHAR+

Le système SYNTHAR+ se présente comme un système de pré-traitements linguistiques agissant directement sur les textes arabes voyellisés pour produire les entrées phonétiques destinées au synthétiseur vocal Multivox. Ce système comporte des traitements morpho-orthographiques et phonologiques modélisés par des règles, des grammaires syntaxiques, un lexique général et des lexiques dédiés. La figure 2 représente l'architecture du système SYNTHAR+.

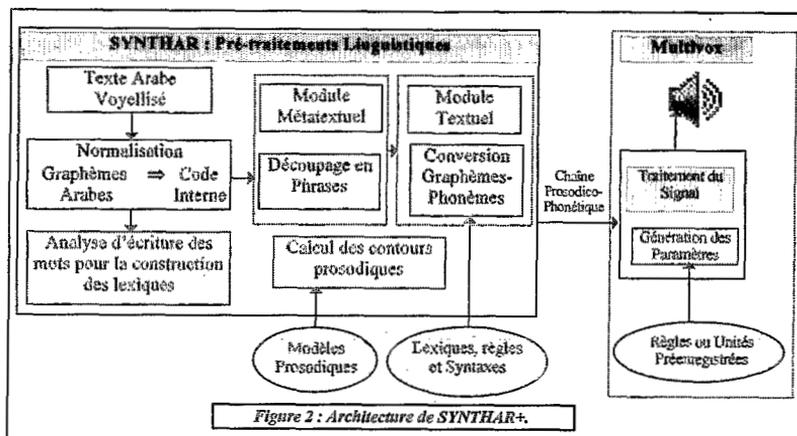


Figure 2 : Architecture de SYNTHAR+.

3.1. Normalisation des caractères et analyse d'écriture des mots

La langue arabe s'écrit de droite à gauche. Elle n'a pas une écriture phonique, car à une phonique peut correspondre plusieurs graphèmes. Par exemple la lettre ع peut prendre quatre formes suivant qu'elle se trouve au début (ع code 183), au milieu (ع code 204), en fin de mot liée (ع code 209) ou en fin de mot isolée (ع code 186). Une représentation interne unique est nécessaire pour faciliter la tâche d'analyse et de construction des règles de transformation graphèmes-phonèmes. La forme isolée est la plus adaptée à cette normalisation [Zemirli 94a].

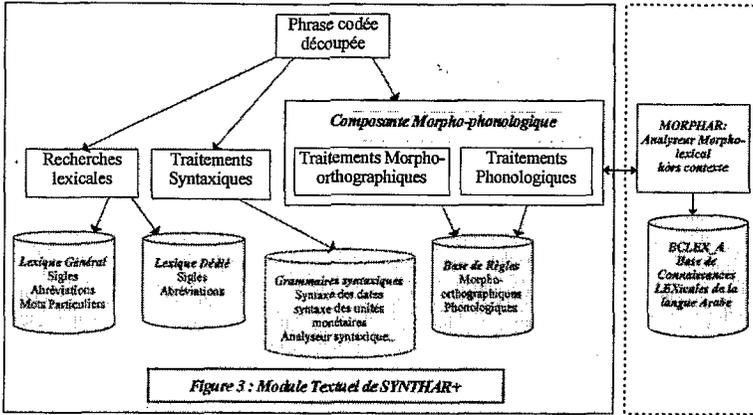
L'écriture des mots, c'est à dire leur formation par voyellisation obéit à certaines règles simples qu'il faudra vérifier. Par exemple deux voyelles ne peuvent se suivre, une voyelle ne peut se trouver en début de mot, une voyelle est manquante, ou un caractère ne pouvant être voyellisé dans une situation particulière [Zemirli 94b]. Ces contrôles sont utilisés pour l'élaboration du lexique général ou des lexiques dédiés des mots particuliers, des sigles et des abréviations.

Le construction d'une base de connaissances lexicales de la langue arabe BCLEX_A et d'un analyseur morphologique MORPHAR [Zemirli 96] seront utilisés pour le développement d'un système de vocalisation automatique de textes arabes.

3.2. Le module textuel de SYNTHAR+

La figure 3 décrit les composants du module textuel de SYNTHAR+ qui a pour rôle de phonétiser les diverses phrases composant un texte voyellisé. Nous considérons qu'une phrase peut comporter des entités comme des sigles, des abréviations, des dates, des notations diverses, etc.

Conformément à la structure générale d'un module de conversion graphèmes-phonèmes [Pérennou 91], [Cotto 92], [Zemirli 95], le module de conversion graphèmes-phonèmes de SYNTHAR+ dispose d'une base de règles, d'une composante lexicale et d'une composante syntaxique. La phrase codée, découpée et présentée à l'entrée du module textuel est le résultat du module de normalisation décrit ci-dessus, et d'un module métatextuel dont l'objet principal est la délimitations des éléments d'une phrase.



3.2.1. La base de règles

Le rôle de la base de règles de notre système est de rendre compte des phénomènes réguliers de la langue arabe. L'ensemble des règles du système de conversion graphèmes-phonèmes est réparti en deux catégories : les règles morpho-orthographiques et les règles phonologiques.

3.2.1.1. Les règles morpho-orthographiques

Les règles morpho-orthographiques ont pour finalité de calculer la représentation phonétique à associer à une représentation orthographique. Leur structure conceptuelle est de la forme :

cg séquence orthographique(s.o) cd → séquence phonétique pour Multivox (s.p)

cg et cd sont respectivement les contextes gauche et droit de la séquence orthographique. Ils peuvent représenter : un séparateur, une consonne, un signe diacritique, etc. Par exemple : 1 indique un séparateur (ponctuation, espace, ...), 2 : indique un caractère différent d'un séparateur, 5 : le signe de la gémination ...

cg	s.o	cd	s.p	Exemples
1	وا	2 ⇒ oo	ذهابوا	ils sont partis
1	ى	2 ⇒ e	رمى	il a jeté
4	ي	0 ⇒ y	يدو	sa main
0	ي	3 ⇒ i	صديقو	son ami
5	س	0 ⇒ ssz	رَسَّام	un peintre
1	;	2 ⇒ in	مَلِك	un roi
1	م	1 ⇒ miim	م	la lettre m
1	ا	2 ⇒ en	عشرون بنتا	vingt filles

- **Consonnes géménées et tenwin**

Les colonnes des tableaux suivants représentent successivement la consonne géminée ou le tenwin (1), l'interprétation de la consonne géminée ou du tenwin (2) et la chaîne phonétique interprétable par le système Multivox (3). Dans le système Multivox, la gémination est considérée comme la réalisation d'une double consonne dont la première est sans vocalisation (soukoun) suivie de la même consonne avec sa voyelle. Cette définition de la gémination devra être contrôlée au cours des expérimentations que nous allons effectuer.

Phonèmes pour Multivoix (3)	Interprétation de la gémignée (2)	gémignée (1)
bbe	بب + ببا	بب
ddi	د + دد	د
sszo	س + سن	س

Dans le cas du tenwin (marque de l'indétermination), il s'agit de la consonne avec sa voyelle suivie d'un "noun" avec un soukoun " (ن).

Phonèmes pour Multivoix (3)	Interprétation du tenwin (2)	tenwin (1)
ton	ت + ن	ت
zen	ز + ن	ز
kin	ك + ن	ك

D'autres règles morpho-orthographiques opèrent en analysant la structure orthographique d'un mot. Ce type de règles morpho-orthographiques est utilisé pour phonétiser correctement des entités comme certains sigles, abréviations, nombres, etc.

• **Les nombres**

Ils peuvent être prononcés dans leur forme isolée chiffre par chiffre ou dans leur prononciation numérique. La première méthode est simple mais présente l'inconvénient de ne pas être très explicite. Le nombre 147 sera prononcé de droite à gauche سبعة - أربعة - واحد. La seconde méthode a l'avantage d'être plus naturelle, les nombres sont prononcés selon leur forme numérale مائة و سبعة و أربعون. Un nombre doit être phonétisé en relation avec la catégorie grammaticale du nom qui lui est associée. Le tableau ci-dessous illustre quelques exemples de phonétisation de nombres.

Chaîne pour Multivoix	Interprétation	Chaîne orthographique
xemzeto rijeelin	خَمْسَةُ رِجَالٍ	5 رجال (5 hommes)
xemszo beneetin	خَمْسُ بَنَاتٍ	5 بنات (5 filles)
celeeceton we e'erbeo:ooone reszmen	ثَلَاثَةٌ وَأَرْبَعُونَ رَسْمًا	43 رسما (43 dessins)
celeecon we e'erbeo:ooone riszeeleten	ثَلَاثٌ وَأَرْبَعُونَ رِسَالَةً	43 رسالة (43 lettres)

• **Les caractères isolés ou sigles inconnus**

Après échec dans la consultation des différents lexiques, les caractères isolés ou sigles inconnus sont générés et prononcés dans leur forme d'épellation standard.

lettre	prononciation
م	m miim
ص	s s'eed
ن	n noon

• **Les abréviations et sigles**

Ils sont prédéfinis dans un dictionnaire qui peut être mis à jour par des modifications, insertions ou suppressions. Toute nouvelle acquisition obéit aux règles d'écriture des mots et de compatibilité des voyelles. Chaque usager pourra créer son propre lexique en fonction de son application.

sigle ou abréviation	valeur dans le dictionnaire
الخ (etc)	إلى آخره .
م.و.ا (I.N.I)	المعهد الوطني للإعلام الآلي
ش.و.س.ح (S.N.C.F)	الشركة الوطنية للسكك الحديدية

Si un mot est constitué uniquement de consonnes et que les accès lexicaux échouent, nous supposons qu'il s'agit d'un sigle non prononçable pour le système de synthèse et par conséquent, nous proposons l'épellation du mot. Dans la nouvelle génération de système de synthèse de textes les connaissances morpho-orthographiques de SYNTHAR+ devraient être utilisées au même titre que les connaissances phonético-phonologiques mises en œuvre dans les modules actuels de transcription graphèmes-phonèmes. Nous rejoignons ici les conclusions de l'utilisation du pré-traitement linguistique TEXOR [Cotto 92] pour le français qui a montré son efficacité pour des textes comprenant de nombreux sigles. D'ailleurs, certaines de ces connaissances morpho-orthographiques sont désormais incluses dans les modules de la nouvelle génération de synthèses (carte PROVERBE ou logiciel SPEECH Engine) multilingues [Rigoni 96].

3.2.1.2. Règles phonologiques : le cas de l'article

Le système Multivox ne dispose pas d'un module phonologique, il faut par conséquent développer les règles nécessaires à la prise en compte des phénomènes phonologiques liés à la langue arabe. Nous pouvons citer à titre d'illustration les transformations phoniques que subit l'article avec l'élision de sa première lettre, sa seconde lettre ou des deux à la fois suivant la nature de la syllabe qui le précède ou de la consonne qui le suit. L'article subit des transformations phoniques selon son contexte immédiat :

1. par rapport à la syllabe qui le précède : le *ا* (*alif*) sera éliminé et le *ل* (*lam*) est prononcé comme la fermeture de la syllabe ouverte à la fin du mot précédent.

Phonèmes pour Multivox	Prononciation	Ecriture
fiilgorfeti (dans la pièce)	فيلغرفة	في الغرفة
febilgorfeti (et dans la pièce)	فيلغرفة	فبالغرفة
qeree'telkiteebe (tu as lu le livre)	قرت الكتاب	قرت الكتاب

2. par rapport au phonème qui suit : le *ل* (*lam*) s'assimilera selon la nature de ce phonème (*solaire ou lunaire*).

Phonèmes pour Multivox	Prononciation	Ecriture
e'ennefiz'eto (la fenêtre)	انفاذة	النافذة

3. l'article peut également subir les deux transformations (1 et 2).

Phonèmes pour Multivox	Prononciation	Ecriture
teh'tee'ennefiz'eti (sous la fenêtre)	تحتنافذة	تحت النافذة

Si pour l'exemple (2) les règles contextuelles élaborées précédemment peuvent être utilisées directement, les cas (1) et (3) ont nécessité l'utilisation de MORPHAR, un analyseur morpho-lexical hors contexte pour la prise en compte du contexte précédent [Zemirli 96]. Cet analyseur procède à un découpage morphologique préalable qui permet de dégager les unités primaires qui composent le mot selon la décomposition suivante :

Postfixe	+	Suffixe	+	Base	+	Préfixe	+	Antéfixe
				غرفة	+	ب + ال	+	ف
هم	+	و-	+	لرصد	+	ي-	+	ف

Les informations du module de segmentation permettent un aiguillage du module d'analyse afin d'effectuer des recherches dans les différents dictionnaires. Par exemple, si un article est détecté alors la base extraite considérée est un nom. Pour le suffixe *ون* la base sera un verbe ou un nom. Pour le suffixe *ات* la base sera un nom ...

Les résultats de l'interprétation fournissent un ensemble d'informations morphologiques associées aux différentes entités primaires formant le mot analysé : classe (nom, verbe ou mot outil), valeur

grammaticale (nom d'agent, adjectif ...), genre (masculin, féminin), nombre (singulier, duel, pluriel), voix (passive, active), aspect (accompli, inaccompli, ...), personne (1ère, 2ème, ou 3ème personne) ...

3.3. La composante lexicale de SYNTHAR+

Contrairement à la base de règles, la composante lexicale d'un système de conversion graphèmes-phonèmes a pour objectif de rendre compte des phénomènes idiosyncrasiques de la langue. En particulier tous les mots dont la phonologie ne peut pas être exprimée correctement à l'aide des règles contenues dans la base de règles doivent faire partie d'un lexique.

La composante lexicale de SYNTHAR+ s'articule autour de deux lexiques :

- le noyau lexical indépendant de toutes applications
- et le lexique propre à l'application.

3.3.1. Le noyau lexical indépendant de toutes applications

Ce noyau comprend l'ensemble des mots qui ne se prononcent pas comme ils s'écrivent. Par exemple, la voyelle longue e " " n'est pas prononcée dans (3). Par contre, elle n'est pas écrite mais prononcée en (2), (3) et (4). Enfin, certains mots s'écrivent d'une façon et se prononcent d'une autre (1).

	<i>Ecriture</i>	<i>Prononciation</i>	<i>Pour Multivox</i>
(1)	مائة	مئة	mie'eto (cent)
(2)	لكن	لكين	leskinne (sauf que)
(3)	هكذا	هكاذ	heekhez'e (comme cela)
(4)	هؤلاء	هواؤلاء	heee'oleee' (ceux-ci)

3.3.2. Le lexique propre à l'application

Le rôle de ce lexique est identique à celui du noyau lexical indépendant de toutes applications. Il est seulement dépendant d'une application. On peut disposer de plusieurs lexiques. Un tel lexique contient les informations nécessaires pour phonétiser correctement certains sigles ou abrégés dépendants d'une application. Dans le cas de l'apprentissage par ordinateur de la langue arabe, on pourra consulter un dictionnaire électronique contenant les différentes abréviations sur les valeurs grammaticales des mots de la langue arabe.

3.4. Traitement des entités respectant une syntaxe spécifique

Pour phonétiser des entités comme les notations de dates ou les unités de mesure précédées d'un nombre, il ne suffit plus de phonétiser les mots dans l'ordre où ils sont exprimés dans le texte. Dans certains cas, il faut procéder à un déplacement de mots. Dans d'autres, il faut ajouter des mots ou des portions de phrases. C'est pourquoi, SYNTHAR+ dispose d'analyseurs syntaxiques dont l'objectif est de déterminer si une séquence de mots représente une entité connue du système. Si l'analyse réussit, une procédure de phonétisation propre à l'entité reconnue est déclenchée. Par exemple, si une séquence de mots respecte la syntaxe *jj/mm/aaaa* ou *jj/mm/aa* le système SYNTHAR+ suppose qu'il s'agit d'une notation de date et par conséquent phonétise la séquence de mots sous une forme propre à l'énonciation de date. La date *15/05/1995* est alors analysée par SYNTHAR+ pour produire la suite phonétique *xemszete o:esere meey e'elf we toszo:o mie'e we xemsze we tiszto:ou* nécessaire au système Multivox. De la même manière, si SYNTHAR+ reconnaît une unité de mesure après un nombre décimal, il phonétise la séquence de mots de telle sorte que le point décimal est remplacé par l'unité de mesure reconnue. *13.5* est alors phonétisé par la suite *celeecete o:esere kiilomitr we nis'f*.

4. Evaluation de l'intelligibilité

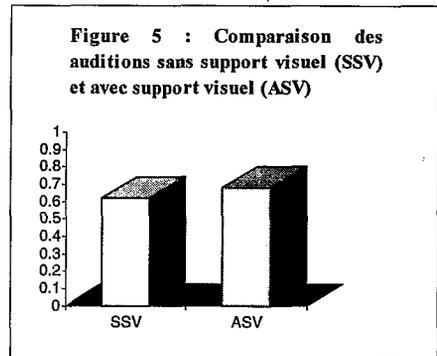
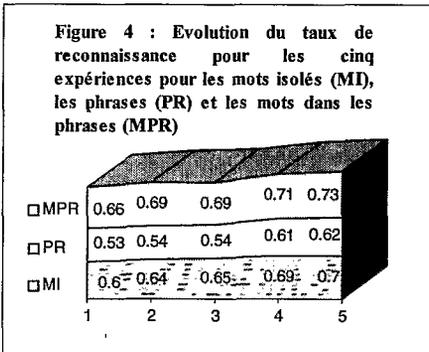
Dans ce qui suit, nous allons évaluer l'apport des pré-traitements linguistiques de SYNTHAR+ pour l'amélioration de l'intelligibilité du système Multivox. Notre objectif est de recenser d'une part, les éventuelles erreurs de phonétisation rencontrées qui pourraient être levées par d'autres pré-traitements et d'autre part, les ambiguïtés phonétiques dues aux système Multivox.

4.1. Présentation du protocole

Les tests d'audition portent sur un corpus de stimuli comportant 40 mots isolés (2 à 5 syllabes) et, 10 phrases comprenant 130 mots au total. Dix auditeurs ayant des niveaux de connaissance différents en langue arabe (3 débutants, 4 moyens et 3 confirmés) ont été soumis aux tests. Les phrases sont courtes (2 à 5 mots), moyennes (6 à 10 mots) ou longues (11 à 20 mots). Elles comportent des nombres, des dates, des abréviations, des sigles, des ponctuations diverses ... Trois auditions sont réalisées sans support visuel (notées 1, 2 et 3 dans la figure 4) suivies de deux auditions avec support visuel (notées 4 et 5). La consigne donnée a été que l'auditeur dise ce qu'il a entendu sans apport d'information supplémentaire.

4.2. Premiers résultats et commentaires

Les graphiques ci-dessous résument les résultats de l'évaluation effectuée. Les auditeurs n'avaient pas écouté de synthèse vocale arabe auparavant. De l'état d'étonnement lors de la première restitution, ils se sont familiarisés avec la synthèse de Multivox. Le taux global de reconnaissance est de l'ordre de 65%. On note cependant une variation des performances : 66% pour les mots isolés (MI), 57% pour les phrases (PR) et 70% pour les mots dans les phrases (MPR). Ces différences peuvent s'expliquer par la nature des stimuli : phrases longues entraînant une perte de sens ou des mots isolés sans contexte. Le support visuel n'améliore pas de façon très significative le score des auditeurs comme le montre la figure 5.



Ce premier protocole d'évaluation de l'intelligibilité de la synthèse vocale produite par Multivox est riche d'enseignements. Nous avons classé les 35% d'échec en deux catégories :

- erreurs dues aux insuffisances du module SYNTHAR+ qui a besoin de connaissances phonologiques supplémentaires sur les mots à synthétiser et des connaissances morpho-syntaxiques sur les phrases à synthétiser
- et les erreurs beaucoup plus nombreuses dues aux insuffisances du système de production et de coarticulation des phonèmes de Multivox.

Les phonèmes ب (b) et د (d) en finale de mot sont interprétés comme un ل (l) ou ن (n). Le ب (b) en position interne est entendu comme un د (d). Nous avons pu grâce à un ralentissement du débit de lecture, restituer de manière plus intelligible ces deux phonèmes. Néanmoins ce ralentissement est opératoire sur le mot ou la phrase, il serait souhaitable que ce paramètre puisse agir sur un phonème ou une suite de phonèmes.

La gémination de la consonne emphatique ط dans le mot نَطَّمَ (il a organisé) provoque les phénomènes suivants : le début du mot se transforme en مَطَّ et la fin du mot en م. Le nouveau mot est donc interprété مَطَّ م (injustice). Nous supposons que ce phénomène est lié à ce que Multivox considère toutes les géménées comme la réalisation de la consonne géminée par une occlusion suivie de la même consonne avec sa voyelle.

5. Conclusion et perspectives

Après la mise en œuvre des premiers pré-traitements lexicaux et phonologiques, nous avons constaté une nette amélioration de l'intelligibilité de la synthèse vocale sous Multivoix. La mise en œuvre des règles contextuelles a permis de prendre en compte dans tous les cas :

- l'assimilation du lam de l'article et de la gémination de la consonne lunaire qui le suit,
- la transformation de l'article par rapport au mot qui le précède.

L'élaboration de lexiques d'exceptions relatifs aux mots particuliers, aux abréviations et aux sigles a permis également de démontrer l'apport bénéfique de la composante lexicale dans le processus de phonétisation.

L'état de mise en œuvre de l'analyseur syntaxique permet de restituer correctement des entités textuelles telles que les dates, les nombres et les entités respectant une syntaxe définie.

Un analyseur morpho-lexical a été développé et nous fournit déjà les valeurs grammaticales hors contexte des mots analysés [Zemirli 96].

Nous procédons actuellement à l'affinement du protocole d'évaluation aussi bien en termes de types que de nombres de stimuli à soumettre à un plus grand nombre d'auditeurs. Par cette étude, nous visons à élaborer des règles phonologiques et/ou morpho-syntaxiques à considérer dans toute conception ou amélioration de systèmes de synthèse de la langue arabe.

Notre prochain objectif est le développement d'un analyseur morpho-syntaxique qui sera utilisé pour la voyellisation des textes et l'amélioration du module de découpage des phrases pour le module textuel.

Nos objectifs à plus long terme telle que l'utilisation de la synthèse arabe dans des tâches de lecture documents électroniques, notamment pour les déficients visuels fait apparaître le besoin d'un système de synthèse de textes intelligible pour toutes les entités du document.

Le développement de la télématique vocale en Europe et les perspectives annoncées pour le monde arabe nous encouragent à poursuivre nos recherches dans le domaine de l'ingénierie linguistique nécessaire à l'essor des technologies vocales, en particulier sur la synthèse à partir de texte.

Bibliographie

- [Aggoun 85] A. Aggoun, "Un système de traitement des connaissances prosodiques pour la synthèse de la parole", Thèse de doctorat, Université de Rennes 1, 1985.
- [Backstrom 89] M. Backstrom, K. Ceder, B. Lyberg "PROPHON - An Interactive Environment for Text-To-Speech Conversion" Eurospeech 89, Paris September 89, Vol 1, pp 144-147.
- [Belhoula 93] K. Belhoula, "A concept for the synthesis of names", Applications of Speech Technology, Lautrach Conference Center, Bavaria, Germany 16-17 Septembre 1993, pp 167-170.
- [Charpentier 89] F. Charpentier, E. Moulines "Nouvelles techniques de synthèse de la parole", l'Echo des Recherches N°137, 1989.
- [Cotto 92] D. Cotto, "Traitement automatique de textes en vue de la synthèse vocale", Thèse de l'U.P.S Toulouse, Décembre 1992.
- [El Shafei 88] A. El Shafei, M.S. Ahmed, A.M. Marzooq "An Arabic Text To Speech System using English allophones" Regional Conference on Informatics and Arabization, pp. 68-85, Tunis 1988.
- [Ghazali 92a] S. Ghazali, Z. Bemmiled, H. Jemini, M. Zrigui, "Synthèse de l'arabe standard à partir de texte par TD_PSOLA : Le traitement des processus phonologiques", Actes 19 ième JEP - Bruxelles, pp. 89-93, 19-22 Mai 1992.
- [Ghazali 92b] S. Ghazali, A. Braham "Voyelles courtes et voyelles brèves en arabe standard : Organisation temporelle", Actes 19 ième JEP Bruxelles, pp.19-24.

SYNTHAR+ : Un Système de pré-traitement de textes arabes en vue de leur synthèse orale sous Multivox

- [Guerti 85] M. Guerti "Contribution à la synthèse de la parole en arabe standard" Thèse Doctorat 3ème cycle CNET 1985.
- [McAllister 89] N. McAllister "The problems of punctuation ambiguity in fully automatic text-to-speech conversion" Eurospeech 89 - European Conference on Speech Communication and Technology - Paris Septembre 1989 - Vol.1 - pp. 538-541.
- [Multivox 94] Multivox : "Standard Arabic Text To Speech System Brochure technique" - AKADIMPEX 1994.
- [Najim 91] M. Najim, P. Baylou, M. Ouadou, A. Radjouani, A. Belhadj "Implementation of the Formant Klatt Synthesiser on the ST18940DSP", Revue Traitement du Signal Volume 9 - N° 2 1991.
- [Oriola 92] B. Oriola, D. Cotto, G. Pérennou, N. Vigouroux, "Consultation de documents électroniques par les personnes handicapées de la vue", Actes des Journées Internationales : L'Interface des Mondes Réels et Virtuels, pp. 489-500, Montpellier, 23-27 Mars 1992.
- [Pérennou 91] G. Pérennou, D. Cotto, M de Calmès, I. Ferrané, J.M. Pécatte, J. Tihoni, "Composantes Phonologiques et Orthographiques de BDEX", Actes des 2èmes Journées Nationales du GRECO-PRC, Communication Homme-Machine EC2 éditeur, pp. 351-362, Toulouse, 29-30 Janvier 1991.
- [Rigoni 96] J. J. Rigoni, "Les applications de la synthèse de la parole dans le monde industriel", Actes des 5èmes Journées Internationales : L'Interface des Mondes Réels et Virtuels, pp. 175-183, Montpellier, 21-24 Mai 1996.
- [Rossi 87] M. Rossi, "Peut-on prédire l'organisation prosodique du langage spontané", Etudes de linguistique appliquée, Aspects Prosodiques de la communication, N° 66, Avril-Juin 1987.
- [Rossi 88] M. Rossi, "Prosodies et technologies vocales" Actes du GRECO-PRC, Paris pp. 62-80, 24-25 Novembre 1988.
- [Rossi 89] M. Rossi, "Ordre, Organisation et Intonation, Mélange de phonétique générale et expérimentale", Publications l'Institut de Phonétique de Strasbourg, pp. 716-733, 1989.
- [Sorin 91] C. Sorin, "Synthèse de la parole à partir de texte" : Etat des recherches et des applications. Actes des 2èmes journées nationales du GRECO-PRC. Communication Homme-Machine. EC2 éditeur, pp. 131-146, Toulouse, 29-30 Janvier 1991.
- [Traber 93] C. Traber, "Syntactic Processing and prosody Control in the SVOX TTS System for German", Eurospeech 93, Berlin Vol 3 - pp. 2099-2102.
- [Zemirli 94a] Z. Zemirli, N. Vigouroux, "Evaluation de la version de démonstration du système de synthèse vocale de la langue arabe Multivox". Rapport Interne de Recherche IRIT - IHM PT - Toulouse Juillet 1994.
- [Zemirli 94b] Z. Zemirli, "Phonétique et morpho-syntaxe de la langue arabe en vue de l'oralisation de textes". Rapport Interne de Recherche N°2 INI Alger Décembre 1994.
- [Zemirli 95] Z. Zemirli, N. Vigouroux, G. Pérennou, "Pré-traitement de textes arabes en vue de leur synthèse orale sous le système Multivox", SNIB'95 - 23 et 24 Oct 1995 - Premier Séminaire National d'Informatique de Biskra - pp 200-210.
- [Zemirli 96] Z. Zemirli, N. Vigouroux, "MORPHAR : Un Analyseur morpho-lexical destiné à l'aide à la construction d'une base de connaissances lexicales de la langue arabe BCLEX_A", EIRA'96 - Colloque International : Langues Situées, Technologie et Communication, Rabat 23-27 Avril 1996, pp 61-64.