

*Académie de Nantes*

**THESE DE DOCTORAT DE L'UNIVERSITE DU MAINE**

Le Mans, France

**Spécialité : INFORMATIQUE**

présentée par

**Sofiane BALOUL**

*pour obtenir le titre de Docteur d'Université*

---

*Développement d'un système automatique de synthèse de la parole à partir du texte arabe standard voyellé*

---

soutenue le 27 mai 2003 devant le jury composé de :

M. Jacques Vergne (GREYC, Caen)	Rapporteur
M. Salem Ghazali (ISLT, Tunis)	Rapporteur
M. Philippe Boula de Mareüil (LIMSI-CNRS, Orsay)	Examineur (Président)
M. Pierre-Yves Le Meur (Elan Speech, Toulouse)	Examineur
M. Marc Baudry (LIUM, Le Mans)	Directeur

## **Résumé**

Le travail de cette thèse est une contribution à l'étude et au développement d'un système de synthèse de la parole à partir du texte arabe standard voyellé basé sur le diphone. Cette contribution intervient à différents niveaux de ce système : construction de la base acoustique, analyse syntaxique, conversion graphème-phonème et génération de la prosodie. L'analyse morpho-syntaxique implémentée repose sur l'utilisation d'un lexique partiel, l'étiquetage par défaut et la propagation de déductions contextuelles. Elle permet le découpage du texte en tronçons (intermédiaires entre le mot et la phrase) non récursifs. L'interface syntaxe-prosodie permet ensuite de distribuer les pauses et de générer les paramètres prosodiques de hauteur et de durée. L'ensemble de ces traitements est intégré dans le système multilingue de synthèse de la parole à partir du texte de la société Elan Speech.

### **Mots clés :**

Synthèse de la parole à partir du texte, traitement de la langue arabe, analyse syntaxique, traitement de la prosodie, langue arabe.

## **ABSTRACT**

The work of this thesis is a contribution to the study and development of a voweled standard Arabic text-to-speech system based on the diphone. This contribution takes place at various levels of this system: construction of the acoustical database, syntax analysis, grapheme-phoneme conversion and generation of the prosody. The morpho-syntactic analysis implemented is based on a partial lexicon, the default tagging and the propagation of contextual deductions. It enables the segmentation of the text into non recursive chunks (intermediaries between the word and the sentence). The syntax-prosody interface enables the allocation of pauses and the generation of the prosodic parameters of pitch and duration. The whole treatments are integrated into the multilingual system of the Elan Speech Company.

### **Key words:**

Text-to-speech, Arabic language processing, syntactic analysis, prosody treatment, Arabic language.

## Remerciements

Je tiens à remercier Monsieur Marc BAUDRY, Professeur au Laboratoire d'Informatique de l'Université du Maine, pour avoir accepté de diriger cette thèse et pour ses précieux conseils et recommandations tout au long de ces années.

Je tiens à remercier Monsieur Jacques VERGNE et Monsieur Salem GHAZALI pour avoir accepté d'être rapporteurs de cette thèse malgré la forte implication et la charge de travail que cela a impliqué.

Je remercie Monsieur Jean-Jacques RIGONI, PDG d'Elan speech, pour m'avoir accueilli au sein de son entreprise et pour m'avoir offert les meilleures conditions pour mener à bien ce travail. En particulier, mes remerciements vont aux membres de l'équipe R&D et à son responsable Monsieur Jacques TOEN pour avoir su partager toute leur expérience et tout leur savoir et pour m'avoir épaulé depuis le début. Toute ma gratitude va à Pierre-Yves LE MEUR et à Philippe BOULA DE MAREÜIL pour leur contribution à la correction et à l'amélioration de ce mémoire et pour leur participation au jury.

Je tiens enfin à remercier les membres de ma famille pour leur incessant soutien et plus particulièrement mes parents qui m'ont guidé sur le chemin des études.

**INTRODUCTION GENERALE ..... 7**

**PARTIE 1 : SYNTHÈSE DE LA PAROLE A PARTIR DU TEXTE ET TRAITEMENT AUTOMATIQUE DE L'ARABE STANDARD ..... 9**

**CHAPITRE 1 : SYNTHÈSE DE LA PAROLE A PARTIR DU TEXTE..... 10**

**1.1. INTRODUCTION ..... 10**  
**1.2. SCHEMA GENERAL D'UN SYSTEME DE SYNTHÈSE A PARTIR DU TEXTE..... 11**  
**1.3. METHODES EN SYNTHÈSE DE LA PAROLE A PARTIR DU TEXTE ..... 12**  
1.3.1. LA SYNTHÈSE PAR REGLES ..... 12  
1.3.2. LA SYNTHÈSE ARTICULATOIRE ..... 12  
1.3.3. LA SYNTHÈSE PAR CONCATENATION D'UNITES ACOUSTIQUES ..... 13  
1.3.4. SYNTHÈSE PAR SÉLECTION DYNAMIQUE D'UNITES NON UNIFORMES A BASE DE CORPUS.. 13  
**1.4. APPLICATIONS DE LA SYNTHÈSE A PARTIR DU TEXTE..... 14**  
**1.5. PRÉSENTATION DU SYSTEME DE SYNTHÈSE D'ELAN SPEECH..... 15**

**CHAPITRE 2 : TRAITEMENT AUTOMATIQUE DE LA LANGUE ARABE..... 19**

**2.1. L'ÉCRITURE ARABE ..... 19**  
2.1.1. LES DIACRITIQUES ..... 19  
2.1.2. LE TANWIN ..... 20  
2.1.3. LA CHADDA ..... 20  
**2.2. PHONÉTIQUE ET PHONOLOGIE DE LA LANGUE ARABE ..... 20**  
2.2.1. LE SYSTÈME VOCALIQUE ..... 20  
2.2.2. LE SYSTÈME CONSONANTIQUE..... 21  
2.2.3. PARTICULARITÉS PHONOLOGIQUES..... 21  
**2.3. PROBLÈME DE LA LANGUE ARABE EN TRAITEMENT AUTOMATIQUE..... 23**  
2.3.1. AGGLUTINATION DES MOTS ..... 24  
2.3.2. VOYELLATION ..... 24  
**2.4. CODAGE INFORMATIQUE DE L'ALPHABET ARABE ..... 26**

**CHAPITRE 3 : CONTEXTE DE L'ÉTUDE..... 29**

**3.1. STRATÉGIES ET RESSOURCES ..... 29**  
**3.2. DIPHONÈME ET LANGUE ARABE ..... 32**  
**3.3. CONSTRUCTION DE LA BASE ACOUSTIQUE ..... 35**  
3.3.1. LISTE DE PHONÈMES ET DIPHONÈMES ..... 35  
3.3.2. ÉLABORATION ET ENREGISTREMENT DES LOGATOMES ..... 36  
3.3.3. SÉGMENTATION ..... 36  
3.3.4. VALIDATION ..... 37

**PARTIE 2 : TRAITEMENTS SYMBOLIQUES ..... 38**

**CHAPITRE 4 : ANALYSE LINGUISTIQUE..... 39**

**4.1. INTRODUCTION ..... 39**  
**4.2. MORPHOLOGIE DE LA LANGUE ARABE ..... 40**  
4.2.1 MECANISME DE DERIVATION ..... 41  
4.2.2. MORPHOLOGIE DU MOT ARABE..... 43  
**4.3. LA GRAMMAIRE EN TRONÇONS : APPLICATION A L'ARABE..... 47**  
**4.4. L'ANALYSE MORPHO-SYNTAXIQUE ..... 49**  
**4.5. TRAVAUX DANS LE DOMAINE ..... 50**

**CHAPITRE 5 : METHODE D'ANALYSE ..... 52**

**5.1. L'ETIQUETAGE MORPHO-SYNTAXIQUE ..... 52**  
**5.2. DESAMBIGUÏSATION..... 61**  
**5.3. PARENTHESES SYNTAXIQUE ..... 61**  
**5.4. ÉVALUATION ..... 62**  
**5.5. DISCUSSION ..... 67**

**CHAPITRE 6 : TRANSCRIPTION ORTHOGRAPHIQUE-PHONÉTIQUE,  
SYLLABATION ET ACCENTUATION..... 68**

**6.1. TRANSCRIPTION ORTHOGRAPHIQUE-PHONETIQUE ..... 68**  
6.1.1. APPROCHES EN TRANSCRIPTION ORTHOGRAPHIQUE-PHONETIQUE ..... 68  
6.1.2. SYSTEMES DE TRANSCRIPTION DE TEXTES ARABES..... 69  
6.1.3. DIFFICULTES EN TRANSCRIPTION GRAPHEME-PHONEME DE L'ARABE ..... 71  
6.1.4. DEVELOPPEMENT D'UN PHONETISEUR POUR L'ARABE STANDARD VOYELLE..... 74  
6.1.5. DISCUSSION ..... 79  
**6.2. SYLLABATION ET ACCENTUATION ..... 80**  
6.2.1. GENERALITES ..... 80  
6.2.2. ÉTUDE DE L'ACCENT ARABE..... 81  
6.2.3. CONCLUSION ..... 84

**PARTIE 3 : PROSODIE..... 85**

**CHAPITRE 7 : ETUDE DE LA PROSODIE..... 86**

**7.1. GENERALITES..... 86**  
**7.2. FONCTION DE LA PROSODIE ..... 87**  
**7.3. LA SUBSTANCE PHONETIQUE DE LA PROSODIE ..... 88**  
7.3.1. LA FREQUENCE FONDAMENTALE ..... 88  
7.3.2. LA DUREE ..... 90  
7.3.3. L'INTENSITE ..... 93  
**7.4. L'INTONATION ..... 93**  
7.4.1. LA DECLINAISON ..... 93  
7.4.2. MODELE DE GENERATION DE L'INTONATION ..... 95  
**7.5. ETUDE DE L'INTONATION ARABE..... 98**  
7.5.1 LE CONTOUR INTONATIF ..... 98  
7.5.2. DISCUSSION ..... 100

**CHAPITRE 8 : ANALYSE ET SYNTHÈSE DE LA PROSODIE..... 101**

**8.1. CORPUS D'ANALYSE ..... 101**  
**8.2. GESTION DES PAUSES ..... 103**  
8.2.1 LES PAUSES ASSOCIÉES AUX SIGNES DE PONCTUATION ..... 104  
8.2.2. LES PAUSES NON ASSOCIÉES AUX SIGNES DE PONCTUATION ..... 104  
8.2.3. DURÉE DES PAUSES..... 108  
8.2.4. LIMITES DU MODÈLE DE GESTION DES PAUSES ..... 109  
**8.3. GÉNÉRATION DE LA FRÉQUENCE FONDAMENTALE..... 111**  
8.3.1. ANALYSE DE L'ACCENT LEXICAL..... 111  
8.3.2. ANALYSE DE L'INTONATION ..... 114  
8.3.3. SYNTHÈSE DE L'INTONATION..... 125  
**8.4. GÉNÉRATION DE LA DURÉE PHONÉMIQUE ..... 128**  
8.4.1. LES VOYELLES ..... 128  
8.4.2. LES CONSONNES ..... 133  
8.4.3. PRÉSENTATION DU MODÈLE ..... 135

**CHAPITRE 3 : ÉVALUATION ..... 136**

**9.1. ÉVALUATION DE L'INTELLIGIBILITÉ ET DU PLACEMENT DES PAUSES..... 137**  
9.1.1. PRÉSENTATION DU PROTOCOLE D'ÉVALUATION ..... 137  
9.1.2. RÉSULTATS..... 137  
9.1.3. INTERPRÉTATION ..... 138  
**9.2. ÉVALUATION DU CONTOUR PROSODIQUE..... 138**  
9.2.1 PROTOCOLE D'ÉVALUATION ..... 138  
9.2.2. RÉSULTATS..... 139  
9.2.3. INTERPRÉTATION DES RÉSULTATS ..... 141  
**CONCLUSION GÉNÉRALE..... 146**

**BIBLIOGRAPHIE ..... 149**

**LISTE DES TABLEAUX ..... 157**

**LISTE DES FIGURES..... 158**

**ANNEXE 1 ..... 159**

**ANNEXE 2 ..... 160**

## Introduction générale

La parole étant le moyen de communication le plus naturel chez l'homme, celui-ci a très vite cherché à l'intégrer dans les interfaces homme-machine. Cela a été rendu possible grâce aux efforts consentis en reconnaissance et en synthèse de la parole. Alors que la première vise à reconnaître les messages de l'utilisateur pour les traduire en action, la seconde a pour objectif de doter l'ordinateur de la capacité à lire des textes à haute voix. Malgré les avancées réalisées ces dernières années dans ces domaines, des progrès restent à faire pour accroître le confort d'utilisation des systèmes actuels.

La synthèse de la parole est un domaine pluridisciplinaire à l'intersection de l'informatique, de la linguistique et du traitement de signal. Hormis Elan Speech, seuls deux industriels à notre connaissance, la société égyptienne Sakhr<sup>1</sup> et l'entreprise belge Babel Technologies<sup>2</sup> commercialisent un système de synthèse à partir du texte arabe, bien que des travaux de laboratoire aient ouvert cette voie depuis plusieurs années. Ceci tient en grande partie à l'insuffisance des ressources linguistiques en traitement automatique de l'arabe et aux caractéristiques intrinsèques de son écriture qui, le plus souvent, est dépourvue de voyelles.

Ce travail se veut une contribution au développement d'un système de synthèse de la parole à partir du texte arabe voyellé à des fins commerciales. Cette contribution intervient à différents niveaux du système : conversion graphème-phonème, analyse syntaxique, génération de la prosodie, dictionnaire de diphtonges, etc. Nous nous plaçons dans un contexte de textes voyellés, l'objectif étant la validation de modèles et d'hypothèses post-voyellation.

De part ses enjeux universitaires et industriels, notre problématique est double :

- D'une part, mener une étude dans le domaine de la synthèse de la parole arabe en proposant des modèles théoriques à différents niveaux de traitement. Notre investigation a essentiellement porté sur l'interface syntaxe-prosodie et sur la modélisation de phénomènes linguistiques en vue de la transcription orthographique-phonétique.
- D'autre part, implémenter les modèles définis et intégrer les connaissances linguistiques arabes dans le système de synthèse multilingue de la société Elan Speech dans le temps qui nous a été imparti. Les stratégies de mise en œuvre adoptées doivent tenir compte des contraintes inhérentes aux systèmes de synthèse vocale et prendre en considération les ressources mises à notre disposition.

Ce mémoire est organisé en trois parties :

---

<sup>1</sup> <http://www.sakhr.com>

<sup>2</sup> <http://www.babeltech.com/>

La première partie est consacrée à la présentation de la synthèse de la parole à partir du texte (SAT) et au traitement automatique de la langue arabe. Nous introduirons la problématique générale en SAT arabe, puis les solutions et ressources adoptées pour y remédier. En particulier, nous donnerons notre point de vue sur la corrélation syntaxe-prosodie en arabe et nous discuterons des contraintes de mise en œuvre dans le cadre d'un système de SAT entièrement automatique.

La deuxième partie décrit les traitements symboliques en SAT arabe. Nous présenterons les principes de notre approche morpho-syntaxique qui segmente la phrase en *tronçons* et les traits linguistiques sur lesquels nous nous sommes appuyés pour sa mise en œuvre. Nous décrirons ensuite les pré-traitements en entrée du système de synthèse, puis nous survolerons les approches utilisées en transcription orthographique-phonétique. Notre système de conversion graphème-phonème et les connaissances linguistiques qui y sont modélisées seront ensuite présentés, suivis des différents points de vue sur la place de l'accent lexical.

La troisième partie présente le modèle de génération de la prosodie et les résultats de son évaluation. Nous proposerons un modèle de prédiction des pauses qui s'appuie sur les signes de ponctuation et sur l'organisation du texte en tronçons. Nous verrons que l'accès à des niveaux linguistiques plus élevés est nécessaire à une meilleure prédiction des pauses. En nous appuyant sur les observations de corpus, nous proposerons deux approches différentes pour la modélisation de l'intonation qui s'appuient sur la structure accentuelle des mots et sur les frontières de tronçons. Nous verrons que le tronçon constitue un meilleur candidat pour le calcul de la ligne de déclinaison. La durée, quant à elle, est modélisée au niveau phonémique par des règles de réduction/allongement en fonction des contextes phonétiques et phonotactiques. En dernier lieu, nous présenterons les résultats de l'évaluation subjective qui a porté sur l'intelligibilité des phonèmes et la place des pauses d'une part et sur le contour prosodique d'autre part.



**PARTIE 1 : Synthèse de la parole à partir du texte et traitement automatique de l'arabe standard**

# CHAPITRE 1 : Synthèse de la parole à partir du texte

Dans ce premier chapitre, nous présenterons le domaine de la synthèse de la parole à partir du texte. Ayant constaté un engouement pour ce domaine au vu des ouvrages et des publications parus récemment, comme le numéro spécial de la revue T.A.L.<sup>3</sup> [Ale01], nous ne nous attarderons pas trop sur les aspects théoriques et renverrons à chaque fois que cela s'avérera nécessaire aux références correspondantes.

## 1.1. Introduction

La synthèse de la parole à partir du texte désigne l'ensemble des traitements permettant à une machine de transformer un texte écrit en message oral. Aucune restriction n'est faite sur la nature des mots à synthétiser (sigle, abréviation, chiffre, date, etc.), ni sur la taille du vocabulaire à traiter.

Le but recherché en SAT est la production d'une voix synthétique qui imite au mieux la voix humaine, tant au niveau de l'intelligibilité des « sons » qu'au niveau du naturel. Cette opération fait appel à des connaissances de natures diverses : informatique (architecture logicielle, temps réel...), linguistique (analyse lexicale, morphologique, syntaxique, sémantique), traitement du signal, etc.

Il faut distinguer la SAT de la synthèse par concaténation de mots. Celle-ci consiste à stocker une suite de mots enregistrés, puis à les utiliser lors de la restitution par la mise bout à bout des signaux correspondant à chaque mot. Cette méthode de « stockage/restitution » est utilisée dans des applications où le vocabulaire est limité et connu à l'avance (horloge parlante, renseignements SNCF, etc.). La qualité de la parole produite par cette méthode est assez bonne par rapport à la SAT, mais son principal défaut est que la mélodie des mots enregistrés séparément ne correspond pas toujours à la mélodie générale des phrases porteuses. De plus, l'ensemble des mots doit être mémorisé au préalable, ce qui nécessite l'enregistrement de nouveaux mots et un temps de mise en œuvre plus important.

Plusieurs systèmes de SAT sont aujourd'hui disponibles. Citons entre autres, les systèmes commerciaux d'Elan Speech, d'AT&T Labs, de Bell Labs et de Babel Technologies ; les systèmes expérimentaux de France Télécom [Big93], du LIMSI<sup>4</sup>, de l'ICP<sup>5</sup> et FIPSVOX [Gau98] développé au LATL<sup>6</sup> de l'Université de Genève. En ce qui concerne la langue arabe, il existe un système expérimental développé à l'IRSIT [Gha92b] et celui d'IBM [Ham00].

---

<sup>3</sup> Traitement Automatique des Langues

<sup>4</sup> Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur

<sup>5</sup> Institut de la Communication Parlée

<sup>6</sup> Laboratoire d'Analyse et de Technologie du Langage

## 1.2. Schéma général d'un système de synthèse à partir du texte

Un système de SAT se compose en général de trois parties (cf. figure 1). Les deux premières parties qui concernent les traitements *de haut niveau* permettent le passage de la représentation orthographique du texte en entrée à une représentation phonétique munie d'une description prosodique. La dernière partie englobe les traitements *de bas niveau* du synthétiseur qui permettent la génération proprement dite du signal acoustique.

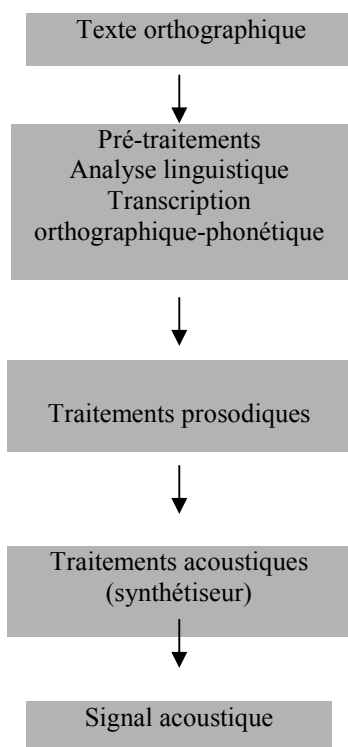


Fig. 1 : Schéma général d'un système de synthèse à partir du texte.

En amont du système, des pré-traitements sont effectués pour le découpage du texte, l'élimination des caractères *parasites* (espaces, sauts de ligne, etc.) et le traitement d'unités spéciales comme les nombres, les abréviations ou les sigles. Une analyse syntaxique plus ou moins élaborée est ensuite appliquée pour l'étiquetage grammatical des mots, suivie d'une transcription orthographique-phonétique (TOP) pour définir la prononciation qui leur est associée. Cette description phonétique accompagnée des informations syntaxiques est passée au module suivant de calcul des paramètres prosodiques liés à l'intonation et au rythme. Les informations calculées par l'analyse linguistique permettent de lever certaines ambiguïtés comme la transcription des nombres en arabe [Zem98a], le traitement des homographes hétérophones en français ou la gestion des pauses et la génération de la prosodie dans différentes langues [Bou97].

Cette première phase de traitements linguistico-prosodiques est du domaine symbolique. Elle fait appel à des modèles théoriques et à diverses connaissances linguistiques **spécifiques** à la langue considérée. À l'inverse, le synthétiseur qui est piloté par des paramètres numériques met en œuvre des techniques en traitement du signal qui sont indépendantes de la langue traitée.

### **1.3. Méthodes en synthèse de la parole à partir du texte**

De la même manière qu'il existe de nombreux procédés pour analyser la parole humaine, il existe différentes méthodes destinées à créer de la parole artificielle. Les méthodes de SAT représentent *le moyen* mis en œuvre pour passer de la représentation symbolique du texte vers le signal acoustique. Elles sont classées en trois ou quatre catégories, selon qu'elles modélisent ou non le fonctionnement de l'appareil vocal.

#### 1.3.1. La synthèse par règles

Cette méthode suppose la connaissance des mécanismes de production et de perception de la parole. 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 ces valeurs cibles est ensuite modélisée à l'aide de règles contextuelles. L'ensemble des valeurs cibles et des règles de transition représente alors les paramètres de commande d'un synthétiseur.

La technique de *la synthèse par formants* est de loin la plus utilisée en synthèse par règles — les formants désignant les maxima de la fonction de transfert (correspondant aux fréquences de résonances) du conduit vocal. Le but de cette technique est de générer un signal sonore à partir des informations sur les formants (fréquences centrales, amplitudes, largeurs de bande) et des règles d'évolution des formants entre phonèmes. Le synthétiseur de Klatt [Kla80] constitue la référence en la matière et a inspiré plusieurs études comme le système de Ferrat [Fer02] pour la langue arabe.

La méthode de synthèse par règles a l'avantage de ne stocker que très peu de données (seulement les valeurs cibles), mais elle a comme principal inconvénient de recourir à un nombre important de règles de transition pour modéliser au mieux les caractéristiques de la parole humaine. La pertinence et le nombre des valeurs cibles est également un élément important, car une représentation inadéquate ou trop simpliste du signal de la parole peut dégrader la qualité globale de la synthèse [Boi00].

#### 1.3.2. La synthèse articulatoire

Cette méthode de synthèse se distingue de la précédente par rapport à l'élément étudié. Alors que la première tente de générer un signal de parole en reproduisant son spectre par exemple, la technique de synthèse articulatoire 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.

### 1.3.3. La synthèse par concaténation d'unités acoustiques

Cette méthode consiste à générer le signal synthétique en concaténant des unités acoustiques obtenues par segmentation du signal naturel. La première idée qui nous vient à l'esprit est d'utiliser le phonème comme unité de base et ceci pour plusieurs raisons : le phonème est une unité bien connue des phonéticiens et son nombre est relativement faible dans les différentes langues. Les expériences ont néanmoins montré que les phases de transition entre phonèmes entraînaient des discontinuités sur le signal reconstitué en raison du phénomène de coarticulation (influence du phonème sur ses voisins). Ceci a amené les chercheurs à choisir des unités couvrant les parties instables du signal et à effectuer les segmentations en des endroits supposés stationnaires comme le milieu du phonème.

Le diphone représente l'unité minimale qui permet de donner une synthèse de bonne qualité. Il s'étale de la partie stable d'un phonème à la partie stable du phonème voisin, couvrant ainsi les phases de transition entre deux phonèmes successifs. L'ensemble des systèmes commerciaux et expérimentaux cités précédemment a adopté le diphone [Eme77] comme unité acoustique. Nous discuterons de la validité du diphone en synthèse de l'arabe dans le chapitre 3 de cette partie.

Cependant, une représentation unique du diphone dans la base acoustique peut être insuffisante à cause des effets de coarticulation dépassant la limite du phonème. Pour pallier ce problème, des unités de taille supérieure au diphone sont proposées, comme la syllabe, la di-syllabe [Che00] (qui s'étale du noyau d'une syllabe jusqu'au noyau de la syllabe suivante) ou des unités de taille variable pour introduire différentes représentations contextuelles des unités dites *sensibles* [Lem96]. Dans cette dernière approche, des unités à trois, quatre ou plusieurs phonèmes (polyphones) sont ajoutées à la base de diphones, la sélection de l'unité appropriée s'effectuant dynamiquement au moment de la synthèse.

### 1.3.4. Synthèse par sélection dynamique d'unités non uniformes à base de corpus

La dernière génération des systèmes de SAT utilise la *synthèse par corpus* [Con99] [Coo00] [Rut00]. Il s'agit de sélectionner des unités de parole de taille variable dans un grand corpus et de les concaténer pour la génération du signal synthétique. Le corpus utilisé peut avoir une taille de plusieurs heures de parole (au moins 2 à 4 heures) offrant pour chacune des unités acoustiques divers contextes phonétiques, phonologiques, syntaxiques, etc. Cette technologie est utilisée dans le système industriel de SAKHR et dans le système expérimental d'IBM [Ham00] pour la synthèse de l'arabe. Elan Speech qui a également développé cette technologie pour la synthèse du français et l'anglais prévoit de l'étendre à la langue arabe.

Nous avons pu voir que les systèmes de SAT utilisent des technologies à base d'unités acoustiques de différentes tailles. D'une manière générale, plus la taille de l'unité acoustique est grande, mieux les phénomènes liés au contexte sont mémorisés et restitués, mais plus la combinatoire des unités à stocker dans la base est importante. Par exemple, une base de diphtonges arabes comptera 1188 unités (44 phonèmes = 28 consonnes + 6 voyelles + 9 allophones + 1 silence), une base de di-syllabes comptera 8832 unités [Che00], tandis que si l'on devait utiliser une base de mots, celle-ci devrait comporter l'ensemble du vocabulaire de la langue, ce qui n'est pas envisageable.

Le choix de la taille de l'unité dépend de l'application visée par le système de synthèse, mais aussi des contraintes liées à son utilisation. Dans le cadre d'un système industriel par exemple, le temps de réponse du système et la qualité de la voix synthétique souhaitée sont autant d'éléments importants à prendre en compte. Une unité large peut accroître la qualité de la voix, mais aura des incidences sur la capacité de stockage et le temps de calcul imparti à la gestion du dictionnaire : chargement du dictionnaire en mémoire, sélection dynamique s'il y a plusieurs unités candidates, etc. Ainsi, la technologie à base de diphtonges est encore utilisée dans les systèmes embarqués à cause des espaces de stockage réduits.

La technique la plus utilisée aujourd'hui pour la concaténation des unités acoustiques, le plus souvent du diphtonge, est la technique PSOLA (Pitch Synchronous Overlap-Add) [Mou90]. Elle permet d'appliquer la description paramétrique de la prosodie, calculée par les modules de haut niveau, aux différents phonèmes, en modifiant leur durée et leur fréquence fondamentale dans le domaine temporel.

Le principe de cette technique est le suivant : le signal original est d'abord transformé en une suite de signaux élémentaires. La longueur des fenêtres est telle que deux signaux élémentaires consécutifs présentent un recouvrement mutuel important. Pour la modification de la durée d'un phonème, des blocs de signaux sont ajoutés ou retranchés entre les signaux déjà existants. La modification de la fréquence fondamentale s'effectue en augmentant ou en réduisant la partie de recouvrement de deux fenêtres consécutives. Cette modification tient compte du caractère voisé ou non de la partie du signal concernée (un son non voisé ne subira pas de transformation puisque sa fréquence fondamentale n'est pas détectée).

Les techniques LPC (Linear Prediction Coding), HNM [Sty96] et MBROLA [Dut96a] [Dut96b] sont également employées pour la concaténation d'unités acoustiques. Cette dernière repose sur le même principe que PSOLA, à la différence qu'elle réalise une compression du signal acoustique (de l'ordre de 10).

#### ***1.4. Applications de la synthèse à partir du texte***

Les applications de la SAT sont très diverses et ciblent aussi bien un public spécifique comme les non-voyants que le grand public. L'article de Sorin [Sor95] et le chapitre de Dutoit

sur la SAT [Boi00] font état de ses principales applications. Nous en exposons ci-dessous les plus courantes, en présentant dans le tableau 1 les applications du système d'Elan Speech :

Telecom	Portails vocaux, lecture de SMS ou d'emails, services de messagerie unifiée, CRM, annuaires et annuaires inverses, web parlant, standard automatique, serveurs vocaux, centres d'appels, etc.
Multimédia	Outils d'aide à la lecture et à l'apprentissage de langues, de relecture, d'assistance aux handicapés, assistant personnel, solution de lecture d'emails ou de fax, web parlant, assistance en ligne, productivité, agents en ligne, etc.
Automotive	Systèmes de navigation embarqué ou déporté, aide à la navigation, systèmes d'alerte et de diagnostic embarqués, info-traffic, lecture d'emails, réservation en ligne, accès Internet, etc.

**Tab. 1 : Applications du système d'Elan Speech.**

- Outils d'aide aux personnes handicapées : le système de SAT joue le rôle de lecteur de textes sur écran ou à distance via un serveur vocal, avec la possibilité pour l'utilisateur de configurer le système, en choisissant la voix de synthèse (masculine ou féminine), le débit de la parole et même la langue souhaitée si le document existe en plusieurs langues.

- Dans les services de lecture vocale : annuaire inverse où l'utilisateur obtient des informations sur une personne à partir de son numéro de téléphone ; consultation des télécopies par voie orale ; lecture vocale de journaux ; consultation de la messagerie écrite, etc.

- Dans le cadre des projets visant le dialogue homme-machine, des prototypes de système de dialogue existent en laboratoire où la synthèse vocale est couplée à un système de reconnaissance et exploite ainsi différentes ressources (informations syntaxiques, sémantiques...) utiles à la génération acoustique du signal.

La SAT connaîtra sans doute de nouvelles applications dans les années à venir, compte tenu de l'attrait grandissant pour les nouvelles technologies. Cela dépendra aussi de la robustesse et de la flexibilité de ces systèmes et de leur capacité à s'intégrer dans des applications manipulant la langue sous ses diverses formes (écrite ou orale). Le principal enjeu des systèmes du futur est de pouvoir dialoguer dans la même langue que l'utilisateur en intégrant dans le signal synthétisé les caractéristiques propres de sa voix, comme le timbre ou encore le rythme.

### **1.5. Présentation du système de synthèse d'Elan Speech**

Le système multilingue de SAT commercialisé par Elan Speech intègre 12 langues : anglais britannique, anglais américain, français, espagnol castillan, espagnol latino-américain,

allemand, russe, portugais brésilien, polonais, italien, néerlandais et récemment, la langue arabe. Il utilise les technologies de synthèse par concaténation de diphtongues (TEMPO<sup>TM</sup>) et par corpus (SAYSO<sup>TM</sup>). Dans la première, il emploie soit la technique TD-PSOLA soit la technique HNM pour la concaténation/modification des unités acoustiques ; dans la seconde, un codeur/décodeur HNM pour le stockage et la restitution des segments de parole.

L'approche adoptée pour le développement de ce système est de type modulaire, l'ensemble de ces modules étant intégré dans une architecture baptisée SYC [Big93]. Le traitement des données se fait linéairement : chaque module reçoit en entrée un flux de données du module précédent, le traite et renvoie le résultat sur le flux de sortie. Les données à traiter ne sont pas obligatoirement homogènes : elles sont organisées en *items* comprenant chacun une identification. Il peut y avoir du texte, du signal de parole, des données partiellement traitées, etc. Les modules sont organisés autour d'un noyau qui régule le fonctionnement général et commande les différentes tâches de la manière suivante (cf. figure 2) :

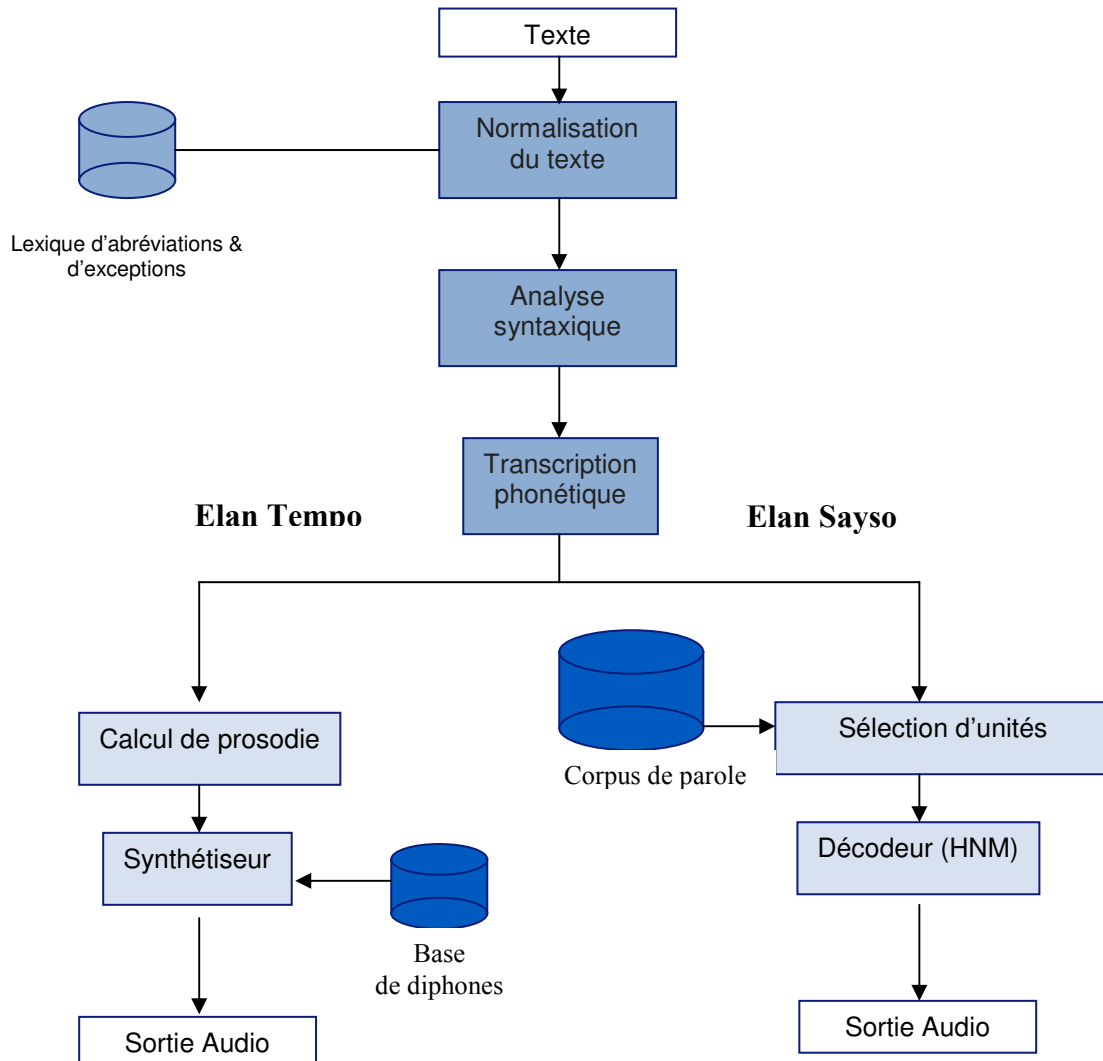
- Normalisation du texte

La mise en forme du texte en entrée se déroule en deux phases : dans la première phase, le texte est découpé en phrases structurées en *mots* (le mot ici est pris au sens large). Ce découpage se base sur des indicateurs de surface comme le blanc, le saut de ligne ou, plus généralement, les signes de ponctuation ; dans la deuxième phase, des pré-traitements sont effectués pour réécrire en toutes lettres les unités spécifiques telles que les nombres, les dates, les sigles ou autres abréviations, etc. Ceux-ci recourent à des lexiques généraux d'exceptions et d'abréviations (pour traiter par exemple des abréviations générales, ex : SNCF) ou à des lexiques spécifiques à l'utilisateur et au domaine d'application d'une part, et à des règles de réécriture pour le traitement des entités non définies dans ces lexiques (ex : nombres) d'autre part. L'ensemble des traitements implémentés à ce niveau est de nature multilingue, seuls les contenus des lexiques et les règles sont propres à la langue traitée.

- Analyse syntaxique

Pour chaque langue, une analyse syntaxique plus ou moins sophistiquée est implémentée pour apporter des solutions aux divers problèmes rencontrés en transcription orthographique-phonétique ou en génération de la prosodie. Le rôle de la syntaxe en SAT arabe sera présenté dans le chapitre 3. Cette analyse s'effectue généralement en trois étapes : analyse morphologique (pour séparer les différents éléments du mot), essentielle pour les langues à forte agglutination comme l'arabe ou l'allemand ; analyse morfo-syntaxique (ou *tagging*) pour l'étiquetage des différents mots en contexte ; parenthésage syntaxique (phrasing ou shallow parsing) pour le découpage du texte en groupes avec éventuellement leur mise en relation.





**Fig. 2 : Architecture du système de SAT d'Elan Speech.**

– Transcription orthographique-phonétique (TOP)

La transcription orthographique-phonétique permet de représenter le texte tel qu'il sera prononcé par le système. La complexité de cette tâche varie selon la langue traitée. Ainsi, la transcription de l'arabe ou de l'espagnol est relativement directe par rapport à celle de la langue française qui présente de nombreuses ambiguïtés de prononciation que seul le contexte syntaxique permet de lever.

L'approche utilisée pour la transcription des différentes langues est de type système expert, hormis l'anglais pour laquelle une approche par analogie est appliquée [Bou01]. La syllabation et l'assignation de l'accent lexical qui détermine la position des syllabes accentuées dans le texte (cf. partie 2, chapitre 6) sont réalisées dans ce module.

Les traitements de bas niveau dépendent de la technologie de synthèse (TEMPO ou SAYSO) employée. Ainsi, dans l'approche à base de diphtonges, un module prosodique calcule les paramètres de hauteur et de durée et les transfère au synthétiseur pour la génération du signal acoustique. Ce dernier extrait la chaîne de diphtonges correspondant à la chaîne phonétique puis lui applique les paramètres numériques à l'aide de l'algorithme TD-PSOLA ou HNM.

Dans l'approche à base de corpus, un module de sélection extrait de la base les unités acoustiques les plus proches de la chaîne phonétique en fonction des coûts cibles (critères linguistiques) et des coûts concaténatifs (critères acoustiques), qui sont des mesures de distance calculées pour chaque unité candidate. Un décodeur HNM restitue les unités d'origine, les concatène puis leur applique un lissage si nécessaire avant la génération du signal acoustique.

## CHAPITRE 2 : Traitement automatique de la langue arabe

Dans le cadre de notre travail de thèse, nous parlerons de la langue arabe en référence à ce qui est communément appelé « l'arabe moderne » ou « l'arabe standard », c'est-à-dire, la langue de communication commune à l'ensemble du monde arabe. Il s'agit de la langue enseignée dans les écoles, donc écrite, mais aussi parlée dans le cadre officiel.

La langue arabe appartient à la famille des langues sémitiques. L'étude de la grammaire arabe a commencé très tôt au milieu du 11<sup>ème</sup> siècle de l'hégire et a donné lieu à d'énormes productions, avant de connaître une période de stagnation qui a duré plusieurs siècles [Boh79]. Ces dernières années, elle connaît un regain d'intérêt, entre autres dans le domaine du traitement automatique.

### 2.1. L'écriture arabe

L'écriture arabe va de droite à gauche et lie les lettres de son alphabet selon des règles de ligatures bien définies, et ceci dans les deux modes manuscrit ou imprimé. Son alphabet compte 28 lettres dont 25 sont des consonnes et 3 (ا و ي) des consonnes ou des voyelles longues selon leur contexte d'apparition (le ي/y/ est une voyelle longue dans فيل/filun/ (« un éléphant ») et consonne dans يَد/yadun/ (« une main »)). La table des symboles utilisés et leurs équivalents en IPA sont présentés en Annexe 1.

La graphie des lettres est différente selon leur position dans le mot. Ainsi, la lettre ع/ε/ est transcrite عَادَ/εAda/ (« il est revenu ») en début de mot, لعبَ/laεiba/ (« il a joué ») en milieu de mot, مَعَ/maεa/ (« avec ») en fin de mot et ودَّعَ/waddaεa/ (« il a quitté ») isolé en fin de mot. Il résulte 78 formes graphiques à partir des 28 lettres. Par ailleurs, la distinction minuscules/majuscules n'existe pas.

#### 2.1.1. Les diacritiques

Les voyelles brèves sont figurées par des symboles appelés *signes diacritiques* (cf. figure 3). Ces symboles sont absents à l'écrit dans la majorité des textes arabes ce qui peut engendrer des ambiguïtés de prononciation dans un système de SAT. L'opération qui consiste à les insérer par une machine dans un texte est appelée *vocalisation automatique* ou *voyellation automatique* (cf. section 2.3.2). Au nombre de trois, ces symboles sont transcrits de la manière suivante :

La *fetha* [a] est symbolisée par un petit trait sur la consonne ( َ/ba/)

La *damma* [u] est symbolisée par un crochet au-dessus de la consonne ( ُ/bu/)

La *kasra* [i] est symbolisée par un petit trait au-dessous de la consonne ( ِ/bi/)

Un petit rond' \_ symbolisant la *soukoun* (سكون) est apposé sur une consonne lorsque celle-ci n'est liée à aucune voyelle (بَعْدَ/baεda/).

### 2.1.2. Le tanwin

Le signe du *tanwin* est ajouté à la fin des mots indéterminés. Il est en relation d'exclusion avec l'article de détermination **ال** placé en début de mot. Les symboles du *tanwin* sont au nombre de trois et sont constitués par dédoublement des signes diacritiques ci-dessus, ce qui se traduit par l'ajout du phonème /n/ au niveau phonétique :

[an] : signe ً (بَّ/ban/)

[un] : signe ُ (بُّ/bun/)

[in] : signe ِ (بِ/bin/)

### 2.1.3. La chadda

Le signe de la *chadda* peut être placé au-dessus de toutes les consonnes en position non-initiale. La consonne qui la reçoit est alors analysée en une séquence de deux consonnes identiques :

Signe ّ /kallama/ (« il a parlé à »).

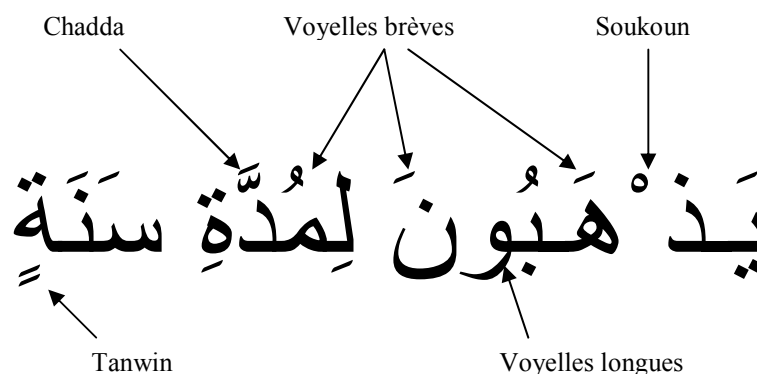


Fig. 3 : Exemple d'une phrase voyellée /yavhabUna limuddati sanatin/  
(« ils partent pour une durée d'un an »)

## 2.2. Phonétique et Phonologie de la langue arabe

El-Ani [Ela70] présente une étude exhaustive des caractéristiques phonétiques et phonologiques de la langue arabe. Nous présentons ci-dessous certaines d'entre elles qui permettront de comprendre nos implémentations futures.

### 2.2.1. Le système vocalique

À côté des trois voyelles brèves /a/ /u/ /i/, il existe trois voyelles longues /A/ /U/ /I/ qui s'opposent aux précédentes par une durée plus importante sur le plan temporel. L'ensemble des voyelles brèves et longues est dit oral car elles sont émises sans l'intervention de la cavité

nasale. Elles sont généralement classées selon le degré d'ouverture du conduit vocal (ouvert /a/, fermé /u/, /i/) et sa position de constriction (/i/ antérieure, /u/ postérieure).

Ces voyelles peuvent avoir des timbres différents selon leur contexte d'apparition :

- Dans un contexte emphatique (au contact des consonnes ص/S/, ض/D/, ط/T/, ظ/Z/), le point d'articulation des voyelles est reporté à l'arrière ;
- Après les consonnes labiales م/m/ et ب/b/, les voyelles sont plus arrondies et se rapprochent du phonème /u/ ;
- Au contact des consonnes ع/ε/ et ه/h/, les voyelles se rapprochent du phonème /a/.

### 2.2.2. Le système consonantique

L'arabe standard contient 28 consonnes qui correspondent chacune à un phonème. La hamza /ʔ/ a un statut particulier en ce sens que certains grammairiens la considèrent comme le 29<sup>ème</sup> phonème : « Lorsque nous commençons un mot par une hamza suivie d'une voyelle (/ʔakala/), nous nous demandons si la première syllabe commence par une voyelle conformément à la transcription phonétique /akala/, ou par une hamza suivie de sa voyelle /ʔakala/ ? » ([Har92], page 62).

À l'instar des autres langues, les consonnes de l'arabe sont classées selon leur mode d'articulation (occlusif, fricatif, nasal, glissant ou liquide), leur lieu d'articulation (labial, dental ou vélo-palatal) et leur voisement (sonore ou sourd). Nous proposons de les grouper en fonction de leurs équivalences dans les autres langues :

- Les phonèmes spécifiques à l'arabe qui n'ont pas d'équivalent dans les langues européennes . ظ/Z/, ط/T/, ض/D/, ص/S/, ح/H/, ء/ʔ/, ق/q/, ع/ε/.
- Les phonèmes qui ont des équivalents dans la langue française : ت/t/, ز/z/, د/d/, س/s/, ش/O/, غ/G/, ك/k/, ج/j/, ف/f/, ب/b/, ل/l/, م/m/, ن/n/, و/w/, ي/y/.
- Les phonèmes qui ont des équivalents dans plusieurs langues telles que l'espagnol, l'allemand ou l'anglais : ر/r/, ذ/v/, ه/h/, خ/x/, ث/c/.

### 2.2.3. Particularités phonologiques

Les caractéristiques phonologiques de l'arabe sont l'emphase, la gémination et le madd.

#### 1. L'emphase

Le mot *emphase* est habituellement utilisé pour rendre compte de manifestations prosodiques liées à l'accentuation volontaire d'une syllabe (cf. partie 3, chapitre 1). Chez les linguistes arabes, il désigne certaines qualités que possèdent les consonnes :

- *L'itbaq* : les consonnes qui ont cette qualité sont ص/S/, ض/D/, ط/T/, ظ/Z/. Celles-ci sont pressées et produites par la langue élevée vers le palais.

- *Le tafhim* : le contraire du *tafhim* est le *tarqiq*. Il traduit une expression acoustique grasse et épaisse de certaines consonnes.
- *L'istila* : cette qualité décrit le mouvement articuloire que fait la langue quand elle meut vers la partie postérieure de la cavité buccale, avec ou sans *tafhim*.

Seules les consonnes /S/, /D/, /T/ et /Z/ possèdent ces trois qualités et sont appelées consonnes *emphatiques* (ou consonnes pharyngalisées). Si nous comparons le français à l'arabe, nous constatons que la différence entre *patte* et *pâte* par exemple est rarement faite en français « standard ». En revanche, cette postériorisation a suscité beaucoup d'intérêt en ce qui concerne l'arabe [Gha81], [Gue87], [Kha99].

Tous ces travaux s'accordent pour dire que les voyelles contribuent à l'emphase. Ainsi, qu'elles soient brèves ou longues, celles-ci possèdent une disposition spectrale différente selon qu'elles sont au contact d'une consonne emphatique ou non-emphatique. Les formants F1 et F2 d'une voyelle dans un contexte emphatique sont plus rapprochés en raison de l'abaissement de F2 (corrélât acoustique de la postériorisation), alors que subsistent des contradictions sur le comportement de F3 [Raj89]. Selon le cas, ceci occasionne un timbre de voyelles différent.

Certaines de ces études affirment que le phénomène de l'emphase dépasse le cadre de la voyelle (ou des voyelles) adjacente(s) et se propage aux phonèmes voisins : « *dans le mot CIVIC2V2... (C=consonne, V=voyelle), si C1 est emphatique, alors la synthèse est plus naturelle quand la propagation de l'emphase arrive jusqu'à C2* » [Gha92b]. En revanche, il existe des divergences sur la portée de cette propagation, en d'autres termes, sur la taille du segment sonore affecté par la consonne emphatique. Une discussion détaillée à ce sujet est présentée par Ghazali [Gha77] [Gha81].

Du fait de sa pertinence au niveau perceptif, la modélisation de l'emphase est primordiale en SAT de l'arabe. Sa prise en compte passe par l'introduction de nouvelles variantes de voyelles dans les contextes emphatiques. Néanmoins, sa mise en œuvre est directement liée à la technique de synthèse utilisée. Rajouani a défini, dans son système à base de règles, un jeu de 6 voyelles brèves et longues emphatisées qui se distinguent des non-emphatisées par la valeur de leurs fréquences formantiques [Raj89]. Dans une approche par diphtonges, Ghazali [Gha92b] et Guerti [Gue87] ont défini des unités acoustiques incluant les variantes emphatisées des voyelles avec l'ensemble des autres phonèmes. Les trajectoires des formants se trouvent ainsi préservées et fidèlement restituées. C'est cette approche que nous avons adoptée dans ce travail (cf. section 3.2).

## 2. La gémination

Au niveau graphique, elle est symbolisée par le signe de la *chadda* qui signifie le doublement de la consonne. Sur le plan phonétique, l'opposition simple/géminée peut se résumer de la manière suivante : pour une consonne non-occlusive, l'opposition se réduit

essentiellement à l'opposition temporelle brève/longue ; pour une occlusive, elle réside au niveau de la durée du silence [Ela70]. Ce rallongement entraîne l'accentuation des propriétés de la consonne (augmentation du caractère emphatique).

Une consonne géminée est un *son* unique pour lequel les organes de phonation ne changent pas de position (les lèvres ne se referment pas après le premier /b/ dans /kab**ba**ra/), d'où la transcription /kab:ara/ qui est plus appropriée [Har92]. Dans beaucoup de langues, ce phénomène permet de mettre en relief un mot dans son contexte, alors qu'il s'avère être un élément distinctif sur les plans morpho-sémantiques en langue arabe [Ela70] : حَضَرَ/ha**D**ara/ (« il a assisté ») est différente de حَضَّرَ/ha**DD**ara/ (« il a préparé ») - la deuxième consonne est géminée.

### 3. Le madd

Ce phénomène concerne l'allongement des voyelles. Il est provoqué par la présence d'une voyelle longue ( /U/, /A/ ou /I/) [Ela70].

La lecture de textes arabes est régie par des règles phonologiques qui ont trait à la contraction des sons, leur élision et à l'assimilation homo-organique des nasales. Certaines de ces règles sont obligatoires, d'autres facultatives ou réservées à certains types de textes, comme le Coran. Nous présentons ci-dessous des définitions brèves de ces phénomènes :

#### – La contraction

Elle est utilisée à cause de la lourdeur de la liaison de deux phonèmes identiques. Elle peut être obligatoire ( قُلْ لهُ /qul/ = قُللَهُ /qullahu/), interdite (dans مَلَلْتُ /malaltu/, le premier /l/ ne doit pas être contracté avec le second /l/) ou permise ( سَرَر /sarara/ = /sarra/);

#### – L'élision

L'élision est le changement qui se produit dans la prononciation du phonème /n/ qui porte une *soukoun* devant certaines consonnes.

#### – L'assimilation homo-organique des nasales

Elle concerne la substitution d'une consonne nasale par une autre consonne. Elle peut se produire à l'intérieur du mot ( أُنْبِتَتْ /?anbatat/= أُمْبِتَتْ /?ambatat/) ou à la frontière de deux mots successifs ( مِنْ بَعْدَ /min baed/ = مِمْبَعْدَ /mimbaed/);

## **2.3. Problème de la langue arabe en traitement automatique**

La langue arabe rencontre deux principaux problèmes en traitement automatique : le premier, général, concerne l'agglutination des mots ; le second, spécifique, a trait à l'absence de voyelles à l'écrit.

### 2.3.1. Agglutination des mots

La plupart des mots arabes sont composés par agglutination d'éléments lexicaux de base (proclitique + base + enclitique). Par exemple, la détermination peut s'exprimer par agglutination de l'article *ال/?al/* avant le mot ( *الولد/?alwaladu/* (« l'enfant ») ou par agglutination d'un pronom personnel après celui-ci ( *وَلَدُهُ/waladuhu/* (« son enfant »). De même, les pronoms personnels peuvent se rattacher aux verbes ( *ضَرَبَهُ/Darabahu/* (« il l'a frappé »), les particules régissant le cas indirect aux noms ( *كَدَارِهِ/kadArihi/* («comme sa maison ») et les conjonctions de coordination aux verbes ( *فَذَهَبَ/favahaba/* («et il est parti »), etc.

Dans toute perspective de traitement automatique, le problème est donc de décomposer le mot en ces différentes parties. Cette décomposition nécessite des connaissances de niveau supérieur en cas d'ambiguïtés (si le mot accepte plusieurs segmentations). Notre analyse morphologique sera présentée dans la partie 2 de ce document.

### 2.3.2. Voyellation

Comme nous l'avons évoqué, les textes arabes sont ordinairement dépourvus de diacritiques. Pour les lire, tout un processus mental est nécessaire : identifier le mot comme appartenant au lexique puis lui attribuer ses voyelles dans son contexte, ce qui nécessite la **compréhension** du texte. Ce problème est similaire à celui de l'accentuation automatique des textes en français, mais dans des proportions beaucoup plus importantes : 28% des mots français en usage sont ambigus contre 95% en arabe (mesures effectuées sur un texte de 23.000 mots) [Deb98].

Pour le traitement automatique d'un texte, il est indispensable d'introduire les voyelles avant (synthèse à partir du texte) ou après (reconnaissance optique) le traitement. Cette opération est appelée *voyellation automatique* ou *vocalisation automatique* quand elle est effectuée par la machine. Elle revient à choisir pour chaque mot du texte une voyellation possible parmi l'ensemble des voyellations acceptées dans la langue. Généralement, elle se déroule en deux étapes :

- Une analyse morphologique qui va assigner à chaque mot non-voyellé l'ensemble des mots voyellés correspondants. Cette analyse recourt à un lexique supposé exhaustif qui intègre toutes les formes canoniques et fléchies des mots. Le tableau 2 donne les résultats de l'analyse morphologiques du mot *كتب/ktb/*.



Verbe accompli	il a écrit	/kataba/	كَتَبَ
Verbe accompli à la forme passive	il a été écrit	/kutiba/	كُتِبَ
Verbe accompli	Il a fait écrire	/kattaba/	كَتَبَ
Verbe accompli	Il a été écrit	/kuttiba/	كُتِبَ
Verbe à l'impératif	Fais écrire	/kattib/	كُتِبْ
Nom au pluriel	Des livres	/kutub/	كُتُبْ
Nom au singulier	Un écrit	/katb/	كُتْبْ
Mot composé du préfixe /k/ + verbe	Comme trancher	/katabba/	كَتَبَ
Mot composé du préfixe /k/ + nom	Comme tranchement	/katab/	كُتِبْ

Tab. 2 : Exemple d'analyse morphologique du mot /ktb/.

- Une analyse syntaxique pour réduire l'ambiguïté au vu du contexte syntaxique. Au mieux, une seule voyellation est retenue. Selon Debili, un texte étiqueté manuellement (100% d'étiquetage correct) présenterait en moyenne 1,4 voyellation par mot, soit 23,5% d'ambiguïtés : « *Ces performances représentent en fait les seuils qui ne pourront jamais être dépassés au sortir de l'étiquetage grammatical* » [Deb98].

- 

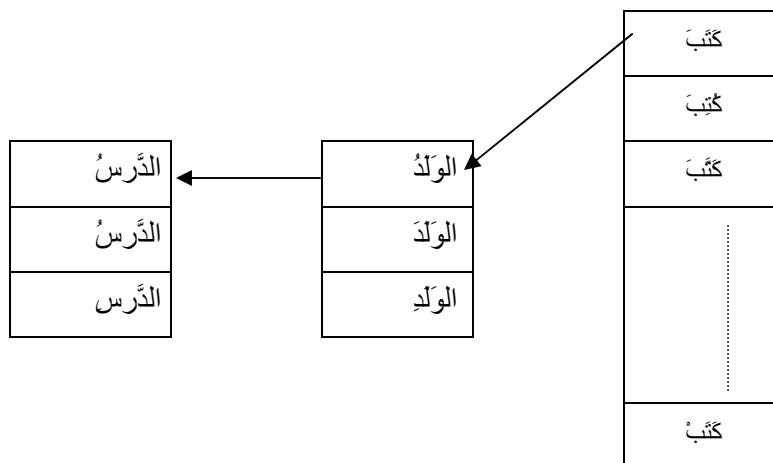


Fig. 4 : Exemple de voyellation de la phrase /katabal waladud darsa/.

La figure 4 représente un exemple de voyellation pour la phrase *كَتَبَ الْوَالِدُ الدَّرْسَ* /katabal waladud darsa/ («l'enfant a écrit son cours») après l'analyse syntaxique. Les mots de cette phrase acceptent respectivement 9 voyellations (/ktb/), 3 voyellations (/ʔlwld/) et 3

voyellations (/ʔddrs/), ce qui engendre  $9*3*3=81$  chemins possibles. Une seule solution est retenue compte tenu de la structure verbe actif + sujet + complément.

- Une analyse sémantique est nécessaire pour réduire l’ambiguïté au vu du *sens* de la phrase. Malheureusement, l’analyse sémantique reste le parent pauvre de la linguistique informatique, mais nous constatons ces dernières années une forte mobilisation scientifique dans ce domaine. En plus de la sémantique, l’analyseur devra comporter des connaissances concernant le contexte dans lequel se déroule le discours (connaissances pragmatiques).

Connaissances sémantiques	قَرَأَتِ الْمُعَلِّمَةُ ←	قَرَأَتِ الْمُعَلِّمَةُ
Connaissances pragmatiques	ضَرَبَ الْوَلَدُ ←	ضَرَبَ الْوَلَدُ

**Tab. 3 : Exemples de voyellation en fonction du sens et du contexte.**

Le tableau 3 représente des exemples de voyellation des phrases قَرَأَتِ الْمُعَلِّمَةُ/qaraʔatil muʕallimatu/ (« l’enseignante a lu ») et ضَرَبَ الْوَلَدُ/Darabal waladu/ (« l’enfant a frappé »). Dans ce deuxième cas, seul le contexte permet de dire si l’enfant a frappé (ضَرَبَ) ou a été frappé (ضُرِبَ).

Il existe très peu de publications scientifiques autour de la voyellation automatique de textes. Nous notons l’article de Debili [Deb98] dans lequel il présente une synthèse de plusieurs années de recherche. Les limites atteintes par son système sont de 68,5% de bonnes résolutions sur les mots. Sa conclusion est la suivante : « Ces performances soulignent l’urgente nécessité d’aller plus loin, vers une analyse reconnaissant et mettant en œuvre des relations syntaxico-sémantiques ».

Par ailleurs, les sociétés égyptiennes SAKHR et française CIMOS<sup>7</sup> ont annoncé depuis peu la commercialisation d’un système de voyellation automatique. Ces sociétés annoncent les scores de 97% (SAKHR) et 95% (CIMOS) de bonne résolution par mot.

## **2.4. Codage informatique de l’alphabet arabe**

L’intérêt que suscite la langue arabe depuis ces dernières années en traitement automatique (traduction, reconnaissance optique et vocale, SAT...) contribue à son intégration dans les systèmes informatiques. De plus, et à l’instar des langues non-latines, elle a bénéficié du développement des nouvelles technologies, en particulier d’Internet, pour répondre aux besoins des utilisateurs dans la gestion et le transfert de documents multilingues.

<sup>7</sup> [www.cimos.com](http://www.cimos.com)

Très tôt s'est posée la question de l'uniformisation du codage des caractères arabes. Le CODAR-U est la première norme de codage arabe proposée par l'ASMO (Arab Standard and Metrology Organization) [Ben97]. Enregistrée à l'ISO sous le numéro 59 en juin 82, elle est devenue successivement la norme CODAR-U/FD puis ASMO 449 réaménagée en ASMO 708. Le tableau 4 représente les normes définies par l'ASMO et retenues par l'ISO<sup>8</sup> comme références internationales.

<b>Norme ASMO</b>	<b>Equivalente ISO</b>	<b>Caractéristiques</b>
CODAR-U/FD	646	- Codage sur 7 bits
449	9036	- Codage sur 7 bits
662	2022	- Codage sur 8 bits - Absence de caractères semi-graphiques - Absence de caractères nationaux
663	2530	- Codage sur 8 bits - Disposition conforme au clavier Qwerty
708	8859/9	- Codage sur 8 bits - Intègre les caractères semi-graphiques et nationaux

**Tab. 4 : Normes de codage ASMO et leurs équivalentes ISO.**

La norme officielle ISO 8859-6 est restée l'apanage des seules plates-formes informatiques utilisant le système UNIX et plus tard, LINUX. Microsoft a défini sa propre « page de code » (CP 1256 – 8 bits) lui permettant de garder en place les accents français dans les polices arabes qu'il diffuse avec Windows. Le Macintosh, quant à lui, respecte à peu près le standard avec son codage MAC-ARABIC.

Les différents codages définis pour l'arabe ont toutefois occasionné quelques problèmes autour des points suivants :

- La représentation de la virgule

Le codage arabe gère deux types de virgules codées différemment : une métrique, identique à la virgule latine et une textuelle inversée et orientée à droite. Cette double définition pose des problèmes au niveau de la représentation des chiffres : 11,15 ou 11•15.

- La manière de saisir les chiffres

Jusqu'à une période très récente, la gestion de l'affichage de la droite vers la gauche n'était possible que dans les systèmes bilingues Arabe-Latin (WIN 95 Arabic Latin). L'insertion des chiffres dans le mode de saisie Arabe (sens droite-gauche) engendre un

<sup>8</sup> International Standardization Organization

décalage vers la gauche au fur et à mesure qu'ils sont entrés (le chiffre 2003 est rédigé 3002).

- La graphie utilisée pour représenter les chiffres

La représentation des chiffres est différente dans la partie orientale des pays arabes (chiffres indiens) et dans les pays du Maghreb (chiffres arabes). Si ces derniers sont bien gérés, les chiffres indiens ne sont pas représentés dans les codages qui n'ont pas déplacé les codes des chiffres latins existants.

En plus de ces quelques problèmes qui sont liés aux pages de code, les logiciels d'édition doivent gérer les différentes formes des caractères en fonction du contexte. Bien que multiforme, chaque caractère a un seul code interne (une seule touche clavier), à l'exception de la hamza /ʔ/ et de la lettre /t/. Pour ces dernières, leurs graphies respectives ا و ؤ ء ة et ة ont des codes différents, ceci en raison de la nécessité de connaissances supplémentaires pour leur transcription (morpho-orthographiques, etc.).

## CHAPITRE 3 : Contexte de l'étude

Nous présenterons dans ce chapitre l'état des recherches en SAT arabe, puis nous dégagerons les stratégies et les ressources adoptées pour le développement de la SAT arabe d'Elan Speech. Par rapport à ces recherches, nous devons prendre en compte les contraintes inhérentes au fonctionnement d'un système industriel multilingue (mémoire de travail, temps réel, etc.) et adapter certains de nos choix en fonction de l'existant (partage des traitements, etc.). Sous un autre angle, nous bénéficions d'une plate-forme logicielle et d'une expérience multilingue qui vont guider notre recherche fondamentale et appliquée.

La complexité des traitements dans une SAT dépend pour une grande part de la langue traitée (une centaine de règles pour la TOP de l'espagnol contre des lexiques d'exceptions et un traitement sophistiqué pour la langue française) et, dans une moindre mesure, du domaine d'application (usage intensif ou occasionnel, tâche précise et limitée ou application à large couverture, etc.). Nous proposons de les classer par rapport à leur niveau de complexité .

### 3.1. Stratégies et ressources

Les systèmes de SAT arabe se sont très peu intéressés à l'intégration des différents niveaux de connaissances linguistiques (phonétiques, phonologiques, morphologiques, syntaxiques...) [Zem98b]. Dans un cadre expérimental, cela n'est pas important en soi. Des phrases dans différentes représentations (orthographiques, phonétiques avec ou sans valeur prosodique) peuvent être soumises au système selon un niveau de travail donné. Dans notre cadre industriel, il est nécessaire d'intégrer ces différents niveaux pour le bon fonctionnement du système. C'est la tâche qui nous est confiée pour l'intégration de la langue arabe dans le système d'Elan Speech.

- Analyse morphologique

Plusieurs analyseurs morphologiques ont été développés pour l'arabe en traitement automatique. Ces analyseurs étaient soit autonomes [Bee01], soit intégrés dans diverses applications comme la voyellation automatique [Deb98], l'analyse syntaxique [Oue02], la traduction automatique [Aze02], la correction automatique de textes [Tai97] ou la transcription orthographique-phonétique [Sar90] [Zem98a].

Le constat que nous avons fait est que très peu de systèmes de synthèse à partir de textes **voyellés** implémentent un traitement morphologique — rappelons que notre système de synthèse suppose ce type d'entrée. Les questions que nous nous sommes alors posées sont les suivantes : Dans quelle mesure ce traitement est-il indispensable en SAT ? La présence de voyelles dispense-t-elle les SAT d'une analyse morphologique ?

La transcription des nombres se fait en examinant les mots auxquels ils sont rattachés. Le genre de celui-ci (masculin ou féminin) détermine leur prononciation (le chiffre 5 est

transcrit /xamsatu/ dans ٥كُتَب (« cinq livres ») et /xamsu/ dans ٥بنت (« cinq filles »). De même, la catégorie grammaticale d'un nombre détermine sa voyelle finale ou sa déclinaison (le chiffre 3 est transcrit /calAci/ dans ٣مراحل (« en trois étapes ») et /calAcu/ dans ٣كان بنات (« il y avait 3 filles ») ; dans le premier cas, il est au cas indirect, dans le second, au cas sujet. Nous reviendrons sur ces problèmes de TOP dans la seconde partie (cf. partie 2 chapitre 6).

Le deuxième point concerne le phénomène d'accentuation, c'est-à-dire, la mise en relief de la syllabe accentuée dans le mot (cf. partie 3, chapitre 3). Pour Rajouani [Raj89] et Bohas [Boh79], l'accentuation doit s'appliquer au mot en excluant ses préfixes et suffixes : ainsi, la première syllabe est accentuée dans /'vahaba/ (« il est parti ») (signe ' avant la syllabe accentuée) et la deuxième dans /fa'vahaba/ (« et il est parti », le préfixe /fa/ est exclu). Ce qui suppose l'emploi d'une analyse morphologique pour la segmentation du mot en préfixe + base + suffixe. De plus, Rajouani a exposé les difficultés suivantes « L'automatisation de la plupart des règles implique le développement d'un analyseur morphologique très puissant associé à un traitement ad-hoc des exceptions, lesquelles demeurent très fréquentes et variées » [Raj89, page 66] (les étiquettes morphologiques ont été insérées manuellement dans son système). Enfin, l'intégration d'une analyse morphologique peut se justifier comme un préalable à un traitement syntaxique dans une SAT.

La première conclusion concerne la nécessité d'une analyse morphologique pour la TOP et l'assignation de l'accent lexical arabe. Cependant, elle doit être adaptée à l'application de la SAT : une analyse morphologique suppose l'exhaustivité des lexiques utilisés (ce n'est pas toujours le cas dans les systèmes expérimentaux qui valident leurs approches en utilisant des sous-lexiques). Nous avons écarté l'approche *lexicaliste* pour deux raisons principales : la première est que très peu de ressources linguistiques sont disponibles ; la deuxième est que la SAT est un domaine à large couverture (même les mots n'appartenant pas à la langue doivent être traités), ce qui nécessite le développement de stratégies qui ne se focalisent pas uniquement sur le lexique.

Nous proposons un traitement morphologique entièrement automatique comme étape préalable au traitement morpho-syntaxique d'une part, et à la TOP et l'assignation de l'accent lexical d'autre part. Ce traitement se base sur des lexiques partiels (lexiques de mots grammaticaux, préfixes, suffixes, etc.) et des traitements de post-validation (règles contextuelles). Il est par contre incontestable que cette restriction engendre des erreurs que seul le lexique pourrait résoudre.

- Rôle de la syntaxe

La contribution de la syntaxe a été mise en évidence dans les SAT traitant de diverses langues indo-européennes [Bac90] [Bou01] [Lib92] [Que92]. Elle peut jouer à deux niveaux : au niveau de la TOP, pour la gestion des homographes hétérophones (mots qui, selon la

catégorie ou le sens, peuvent être prononcés de façon différente), ou au niveau de la génération de l'intonation et des pauses.

En ce qui concerne la langue arabe (nous traitons ici de l'arabe standard), peu de recherches ont abordé cette question, et les avis se rapportant au rôle de la syntaxe sont divergents. Les premières études affirment qu'il existe une relation privilégiée entre la prosodie (les maxima du contour intonatif) et la syntaxe [Ess88] [Raj89] : elles supposent une analyse syntaxique sophistiquée et un générateur prosodique fondé sur la structure syntaxique ainsi produite. Mais en l'absence d'une analyse syntaxique automatique, l'étiquetage des mots en parties du discours se fait manuellement [Elk90], ce qui est inenvisageable dans le cadre d'un système automatique de SAT.

Des recherches plus récentes sur la prosodie arabe réfutent cette nécessité d'un traitement syntaxique, et suggèrent que la prosodie peut être générée indépendamment, sur la base de critères phonétiques, phonologiques et phonotactiques [Saf01], bien que certaines opérations soient interdites à certains endroits comme après la préposition في /fi/ (insertion d'une pause par exemple).

Notre objectif ici est de produire une analyse syntaxique en vue de la synthèse vocale. L'analyse syntaxique n'est pas un but en soi, mais elle doit être guidée par les contraintes inhérentes au système de synthèse : souplesse, robustesse (pour des applications à large couverture), rapidité (temps réel) et qualité globale acceptable. Pour répondre à ces exigences, il n'est pas souhaitable, par exemple, de rendre compte de la grammaticalité des phrases à synthétiser (en rejetant les phrases n'appartenant pas à la langue) : le système doit traiter n'importe quelle phrase en entrée. De plus, il n'est pas nécessaire d'explorer pour chaque phrase l'ensemble des solutions possibles : l'analyse syntaxique doit être déterministe.

C'est dans ce contexte que nous proposons une démarche qui se distingue des deux tendances précédentes, et qui prône une position intermédiaire : nous verrons d'abord que la syntaxe est incontournable, mais qu'une analyse syntaxique superficielle, partielle (*shallow/partial parsing*) peut suffire au calcul de la prosodie (au moins pour établir une bipartition entre mots clitiques et non-clitiques). Cette analyse reprend les principes de Vergne [Ver99] ; ensuite, ce traitement syntaxique est entièrement automatique, à supposer que le texte en entrée soit voyellé. C'est peut être la principale contribution de notre travail.

- Prosodie

La prosodie arabe en SAT est un champ de travail encore mal exploité. L'un des premiers modèles de génération de la prosodie apparu dans la littérature est le MIR (Modèle de l'Intonation de Rabat), développé à la Faculté des Sciences de Rabat [Ess88] [Raj89] [Elk90]. Celui-ci se fonde sur des marqueurs syntaxiques introduits manuellement et

s'applique à des phrases simples (Verbe + Sujet + complément, Nom + Attribut). Nous reviendrons sur les caractéristiques du MIR et ses limites dans la partie 3 chapitre 7.

Nous proposons dans ce travail un modèle de génération prosodique entièrement automatique qui se fonde sur les informations morpho-syntaxiques fournies par les modules en amont. Ce modèle s'applique à l'ensemble des phrases de la langue arabe, en rendant compte de phénomènes presque universels comme la déclinaison ou la collision d'accents.

- Transcription orthographique-phonétique

La TOP à partir de textes voyellés est une tâche bien maîtrisée en traitement automatique de la langue arabe du fait de la forme des graphèmes qui est proche de la forme des phonèmes. Il existe différents systèmes de conversion graphème / phonème [Sar90] [Gha92b] [Zem98a] qui se distinguent par le niveau de connaissances qu'ils mettent en œuvre. Mais très peu de ces systèmes se sont intéressés aux outils et aux formalismes informatiques pour l'implémentation des règles de conversion.

Nous proposons dans ce travail un système de conversion graphème/phonème développé à l'aide de `flex`, un générateur d'analyseur lexical, qui facilite les tâches de gestion et de mise à jour de la quantité de règles implémentées. Cet outil ainsi que les connaissances mises en œuvre seront présentés dans la partie 2, chapitre 3.

- Gestion des pauses

Aucun travail à notre connaissance ne traite de la prédiction automatique des pauses en SAT arabe, en référence aux pauses qui ne sont pas associées aux marques de ponctuation. La prise en compte de ces marques pour la gestion des pauses constitue par ailleurs le traitement minimal pour une SAT. Nous proposons un modèle de prédiction de la position et de la durée des pauses qui s'appuie sur les frontières syntaxiques et sur des considérations phonotactiques rendant compte des mécanismes de phonation.

La première étape de notre travail a été la fabrication d'un dictionnaire de diphtonges avec le concours d'un expert phonéticien. Celui-ci a défini la liste de diphtonges à prendre en compte et le corpus de logatomes à partir duquel ils devaient être extraits. Nous avons pour notre part segmenté ces logatomes, mis en forme la base de diphtonges générée et intégré cette base dans le système de synthèse d'Elan Speech. Avant de décrire les différentes étapes de fabrication du dictionnaire, nous présenterons l'état des recherches par rapport à l'utilisation du diphtongue en arabe.

### **3.2. Diphtongue et langue arabe**

Les premières études à avoir utilisé le diphtongue arabe comme unité de base ont été menées respectivement à l'ENPA<sup>9</sup> en collaboration avec le CNET<sup>10</sup> [Gue83] et à la Faculté

---

<sup>9</sup> École Nationale Polytechnique d'Alger



des Sciences de Rabat [Mou84]. La technique LPC a été d'abord utilisée pour la concaténation/codification des unités acoustiques. Elle a été supplantée par la technique PSOLA [Mou90] dans d'autres systèmes de SAT [Mou87] [Gha92b] et plus récemment dans des systèmes intégrant le moteur de synthèse d'Elan Speech [Zak00a] [Saf01] [Bal02a].

Mais le diphone est-il adapté à la SAT arabe ? Rappelons que le diphone va du milieu d'un phonème au milieu du phonème voisin, couvrant ainsi les phases de transition [Eme77]. Le principe de cette approche est le suivant : mémoriser puis restituer la phase instable entre phonèmes plutôt que de les modéliser. Le point de concaténation sera le milieu du phonème supposé stable. Exemple : le mot مَدْرَسَةٌ/madrasatun/ (« école ») sera restitué à partir de la séquence de diphones /#m/ /ma/ /ad/ /dr/ /rs/ /sa/ /at/ /tu/ /un/ /n#/ . Les diphones /#m/ et /n#/ représentent respectivement un segment de silence suivi de la première partie du phonème /m/ et la deuxième partie du phonème /n/ suivie d'un segment de silence.

Cependant, le choix du diphone peut poser quelques problèmes en raison du phénomène de l'emphase. Examinons le mot وَصَفَ/waSafa/ (« il a décrit ») dont la séquence correspondante est /#w/ /wa/ /aS/ /Sa/ /af/ /fa/ /a#/ . Pour beaucoup d'auteurs, les voyelles au contact de la consonne emphatique (avant et après) constituent le segment minimum affecté par l'emphase [Gha92a]. La consonne ص/S/ est emphatique, donc, la première partie de la voyelle /a/ du diphone /Sa/ est emphatique. Par contre, sa deuxième partie qui se trouve dans /af/ ne l'est pas forcément si ce diphone n'a pas été extrait d'un contexte emphatique. D'où, la nécessité d'introduire des variantes emphatisées pour les différentes voyelles. Ceci est également valable pour la voyelle précédant le ص/S/. La chaîne sera alors /wâ/ /âS/ /Sâ/ /âf/ /fa/ /a#/ , ou /â/ est la variante emphatisée de /a/.

Mais qu'en est-il de la propagation de l'emphase qui dépasse la frontière du phonème ? Pour illustrer ce phénomène, nous présentons deux points de vue différents, ceux de Ghazali [Gha77] et Rajouani [Raj89] qui respectivement ont dressé une étude détaillée à ce sujet.

Pour Ghazali, la propagation de l'emphase est liée au substrat dialectal. En ce qui concerne le dialecte tunisien, l'emphase est arrêtée dans le sens régressif par la frontière du mot, et dans le sens progressif par une voyelle palatale. Rajouani quant à lui propose les règles suivantes sur un substrat marocain :

- Le domaine minimal est CV ou VC ;
- Dans le sens régressif, l'emphase se propage sur toutes les voyelles du mot ;
- Dans le sens progressif, l'emphase est arrêtée par les voyelles /u/ /U/ /i/ /I/ sans dépasser la frontière du mot ;

---

<sup>10</sup> Centre National d'Études des Télécommunications

- Une consonne emphatique placée à l'avant-dernière position d'un mot affecte la dernière voyelle quel que soit son contexte phonétique.

Les deux auteurs sont d'accord sur le fait que le domaine de l'emphase est toujours le mot et que la propagation affecte l'ensemble des voyelles dans le sens régressif. Par contre, ils divergent sur les éléments atténuant la propagation dans le sens progressif. La propagation de l'emphase dans leurs systèmes à base de diphtonges a été modélisée au niveau phonologique. La modélisation de l'emphase reste une tâche difficile : « *Bien que complexes, ces règles ne constituent pas un modèle très fidèle de comportement de ce phénomène de coarticulation, à cause notamment de la nature non binaire de ce trait qui est présent à des degrés divers en fonction de la distance séparant un segment affecté de la consonne emphatique source.* » [Gha92b]. En ce qui nous concerne, nous avons modélisé l'emphase au niveau du segment minimal VC et CV dans notre système actuel à base de diphtonges. Dans la perspective d'étendre la technologie SAYSO à la langue arabe, nous envisageons de modéliser l'emphase au niveau phonétique compte tenu de la variété des unités acoustiques dans le corpus.

En plus de la limite précédente, Mouradi [Mou87] a mené une étude sur la validité du diphtonge en arabe et a énuméré les limites suivantes :

- Les plosives géminées sont mal perçues lorsqu'elles sont réalisées avec le *burst* d'une plosive simple. Exemple : supposons que le diphtonge /ka/ ait été extrait du logatome ركب /rakaba/ (« il est monté ») et utilisé pour la génération du mot رَكَّزَ /rakkaza/ (« il a fixé ») avec la chaîne /#r/ /ra/ /ak/ /kk/ /ka/ /az/ /za/ /a#/ . Dans ce cas, la consonne géminée de /rakkaza/ intègre le *burst* de son équivalente simple de /rakaba/.
- Certaines consonnes dépendant fortement du contexte requièrent plusieurs occurrences dans la base. Exemple : si le diphtonge /#ε/ est extrait du logatome /εalamun/ عَلَمٌ (« drapeau ») et utilisé pour générer la séquence /εilmun/ عِلْمٌ (« science »), alors le phonème /ε/ est mal perçu. Il conviendrait donc de définir deux occurrences ou plus du diphtonge /#ε/ dans des contextes différents.
- Dans le même ordre, certaines consonnes sont voisées/non voisées selon leur contexte phonétique. Ainsi, le phonème laryngal \*h/ est voisé dans un contexte intervocalique ذَهَبَ /vahaβa/ (« il est parti ») et non voisé en début de mot هَرَبَ /haraba/ (« il s'est enfui »). Il est préférable selon Mouradi d'enregistrer la version voisée de ces consonnes et de proposer des solutions ad hoc au moment de la restitution s'ils se trouvent dans des contextes non-voisés.

Enfin, l'utilisation du diphtonge peut engendrer des discontinuités temporelles au point de concaténation consonne/voyelle. Ces discontinuités sont moins perceptibles entre deux noyaux vocaliques voyelle/voyelle [Che00].

Comme le montrent ces exemples, le choix du diphone comme unité acoustique présente quelques limites, bien qu'il produise globalement une meilleure qualité par rapport à la modélisation par règles. Pour pallier ces limites, le système PARADIS<sup>11</sup> [Che00] utilise la di-syllabe (elle couvre le segment s'étalant entre deux voyelles couvrant un nombre quelconque de consonnes) qui a l'avantage d'atténuer le problème de discontinuité temporelle et de résoudre les problèmes décrits par Mouradi. Cependant, cette approche ne fait que retarder le problème de l'emphase qui peut s'étendre au-delà de la di-syllabe.

L'approche par corpus est sans doute la plus appropriée à l'heure actuelle pour la résolution des problèmes posés dans les SAT à base de concaténation d'unités acoustiques. Mais cela dépend aussi de la pertinence des algorithmes de sélection dynamique et de la couverture phonétique des corpus. Ainsi, si le mot وَصَفَ/waSafa/ (« il a décrit ») ne se trouve pas dans le corpus, les algorithmes doivent extraire les unités acoustiques /#w/ /wa/ /aS/ /Sa/ /af/ /fa/ /a#/ dont les contextes phonétiques sont les plus proches de /waSafa/ (par exemple /#w/ /wa/ /aS/ de وَصَلَ/waSala/ (« il est arrivé ») et /Sa/ /af/ /fa/ /a#/ de قَصَفَ/qaSafa/ (« il a tiré »)).

### 3.3. Construction de la base acoustique

La fabrication d'une voix de synthèse est une étape importante dans le développement d'une SAT, car une voix de qualité médiocre dégrade d'autant la qualité globale du système. En général, le choix d'une voix se fait selon un protocole bien défini. D'abord, des locuteurs présélectionnés (de 3 à 5) enregistrent leur voix dans des conditions de prise de *son* optimales (chambre sourde, avec un micro de bonne qualité). Ensuite, une évaluation permet de désigner la voix finale.

#### 3.3.1. Liste de phonèmes et diphtongues

L'expertise phonétique a permis de recenser l'ensemble des phonèmes de la langue arabe, puis l'ensemble des diphtongues compte tenu des séquences interdites (par exemple, écarter les séquences voyelle/voyelle). Ainsi, en plus des phonèmes de base qui correspondent aux 28 consonnes, aux 3 voyelles brèves et aux 3 voyelles longues, les allophones suivants ont été ajoutés :

- 6 allophones /â/ /û/ /î/ /Ā/ /Ū/ /Î/ pour rendre compte des contextes emphatiques. Ils correspondent aux versions emphatisées des phonèmes /a/ /u/ /i/ /A/ /U/ /I/ ;
- 2 allophones /ũ/ et /ĩ/ pour rendre compte des variations des phonèmes /u/ et /i/ dans les contextes non-emphatiques *voyelle/consonne/silence* (en fin de mot) ;

<sup>11</sup> Psola Arabic Di-syllable concatenation based System.

- 1 allophone /Ú/ pour rendre compte des variations du phonème /U/ dans certains mots d'emprunt (mots étrangers). Exemple : *تكنولوجيا*/tiknÚUjijap/ (« technologie »).

- Le pseudo-phonème # correspondant au segment de *silence* est introduit pour rendre compte des diphtonges en position initiale et finale. Exemple : *مَنْ*/#man#/ (« qui »).

Au total, nous avons obtenu 44 phonèmes : 28 phonèmes (consonnes) + 6 phonèmes (voyelles non emphatisées) + 6 allophones (voyelles emphatisées) + 1 silence (phonème #) + 2 allophones (/u/ et /i/ dans le contexte VC#) + 1 allophone (/U/ dans les mots d'emprunt) .

Nous avons ensuite établi la liste des diphtonges en tenant compte des réalisations attestées dans la langue. Ainsi, les paires *silence/voyelle*, *voyelle/voyelle*, *voyelle non emphatisée/consonne emphatique* et *consonne emphatique/voyelle non emphatisée* ont été écartées, ce qui donne un total de 1188 diphtonges.

### 3.3.2. Elaboration et enregistrement des logatomes

Nous avons ensuite élaboré une base de logatomes, des mots dépourvus de sens incluant chacun un seul diphtonge dans un contexte phonétique neutre, afin de minimiser les effets de coarticulation. Ces diphtonges se trouvent dans une syllabe non accentuée. Par exemple, le diphtonge /ba/ est inclus dans le logatome /tabata/. Les diphtonges ont l'une des structures suivantes :

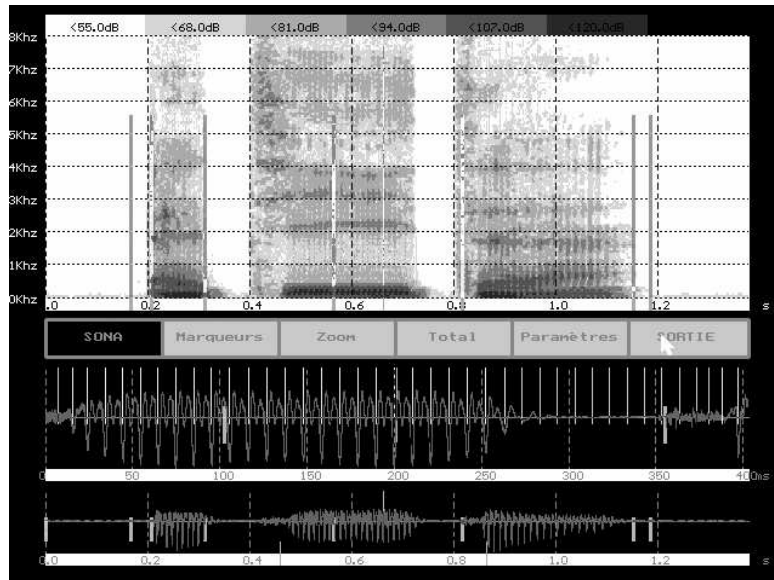
1. CV (C = consonne, V = voyelle) avec C non emphatique, emphatique ou *emphatisée* par extension de l'emphase (c'est à dire que C est affectée par l'emphase mais n'appartient pas à la liste *ص ظ ط*) ; V brève ou longue, emphatisée ou non emphatisée selon C ;
2. VC avec les mêmes contraintes phonétiques que CV ;
3. V# pour les voyelles en fin de mot, V étant emphatisée ou non ;
4. C1C2 avec C1 et C2 identiques ou différentes ;
5. #C et C# pour les consonnes en début ou en fin de mot ;

L'ensemble des logatomes a été enregistré par un locuteur ayant une bonne élocution de la langue arabe, sans accent dialectal marqué.

### 3.3.3. Segmentation

La segmentation consiste à extraire les diphtonges des différents logatomes enregistrés. Cette segmentation a été effectuée manuellement à l'aide de *Sonalog*, un outil de segmentation de la société Elan Speech. Son interface présente le sonogramme (en haut), le signal acoustique (au milieu) et un zoom de la partie sélectionnée (en bas). Pour extraire le diphtonge /ba/ du logatome /tabada/ par exemple, il faut le délimiter en plaçant deux marques

de segmentation (les lignes verticales sur le sonogramme) : la première au milieu du phonème /b/ et la seconde au milieu du phonème /a/. Une troisième marque de repère doit être placée au milieu du diphone (cf. figure 5).



**Fig. 5 : Interface de l'outil Sonalog.**

### 3.3.4. Validation

La validation a consisté à écouter l'ensemble des unités extraites, ce qui a nécessité certaines retouches pour les dipphones mal perçus. Une fois ces corrections effectuées, la base de dipphones a été mise en forme pour son exploitation par le système de SAT. La taille de la base obtenue est de 3,5 Mégabits à la fréquence d'échantillonnage de 16 kHz et la résolution de 8 bits par échantillon (214 secondes).

## **PARTIE 2 : TRAITEMENTS SYMBOLIQUES**

# CHAPITRE 4 : Analyse linguistique

## 4.1. Introduction

Le rôle de l'analyse syntaxique est de reconnaître la structure de la phrase : identification du sujet et des objets liés au verbe, reconnaissance des groupes, etc. [Sab89]. Les connaissances syntaxiques, appelées aussi *grammaire*, expriment les relations entre les catégories de mots. Leur contenu indiquera que tel ou tel autre phénomène syntaxique sera examiné, sachant qu'il n'existe pas d'analyseur syntaxique rendant compte de l'ensemble des phénomènes syntaxiques d'une langue. Traditionnellement, l'analyse syntaxique se déroule en deux phases (cf. figure 6) :

- Une analyse morpho-lexicale qui assigne à chaque *token* (mot au sens large) un ensemble d'étiquettes morphologiques hors contexte. Cette analyse consulte généralement des ressources lexicales supposées exhaustives ;
- Un traitement syntaxique qui fournit l'ensemble des structures acceptables de la phrase. La grammaire utilisée est supposée couvrir l'ensemble des structures syntaxiques appartenant à la langue.

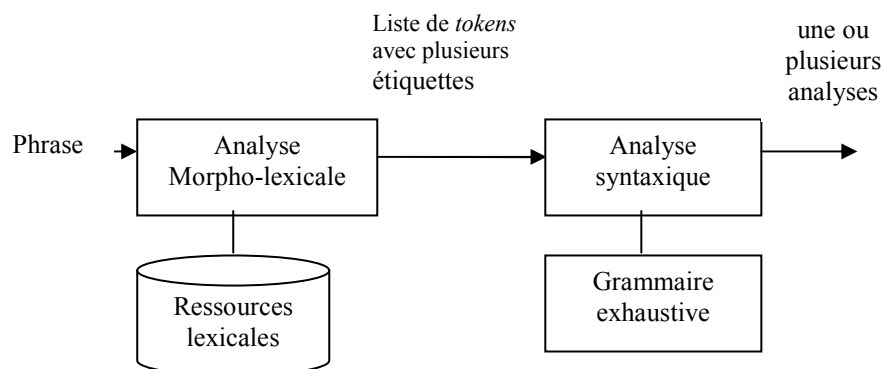


Fig. 6 : Analyse syntaxique traditionnelle.

Le problème rencontré dans les analyseurs syntaxiques traditionnels est celui de la *combinatoire*, qui peut être d'origine lexicale (plusieurs étiquettes pour un *token*) ou structurale (plusieurs structures pour une phrase). La complexité de cette double ambiguïté suit une fonction exponentielle qui augmente avec le nombre de mots dans la phrase [Ver98a]. L'implémentation d'une telle analyse est de ce fait inconcevable dans des systèmes opérant en temps réel, compte tenu de l'incidence que peut avoir cette combinatoire sur le temps de calcul.

De nouvelles méthodes ont été développées pour pallier ce problème de combinatoire. Fondées sur l'étiquetage morpho-syntaxique ou *tagging*, ces méthodes permettent de réduire l'ambiguïté lexicale en introduisant des informations sur le contexte immédiat des mots (cf. figure 7). Le *tagger* vient ainsi se substituer à l'analyseur morpho-lexical avec comme

nouvelle ressource une base de connaissances contextuelles. Ces connaissances peuvent être soit de type probabiliste, si le tagger utilise des informations statistiques sur la contiguïté des mots, soit sous forme de règles explicites.

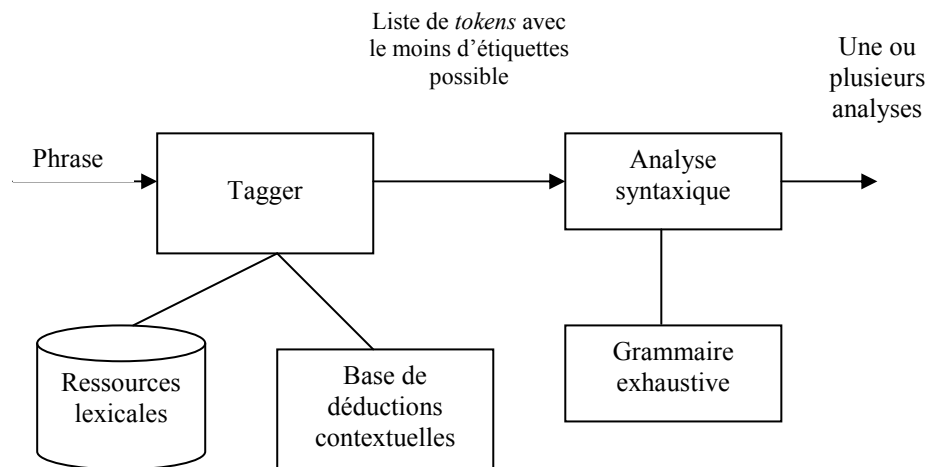


Fig. 7 : Analyse syntaxique à base de *tagger*.

L'analyse morphologique constitue l'étape préalable à tout traitement syntaxique, qu'il soit fondé sur un analyseur morpho-lexical ou sur un tagger. La section suivante présente les bases de cette analyse et son application à la langue arabe.

## 4.2. Morphologie de la langue arabe

En arabe, l'analyse morphologique est d'autant plus importante que les mots sont fortement agglutinés, c'est-à-dire qu'ils sont formés dans leur majorité par assemblage d'unités lexicales élémentaires. Le rôle de l'analyse morphologique est alors de :

- Découper les mots en unités lexicales élémentaires ;
- Attribuer à chaque unité une ou plusieurs valeurs morphologiques.

Ce traitement morphologique nécessite que les unités à extraire soient connues et répertoriées dans des lexiques. La démarche à suivre consiste donc à segmenter les mots de manière à ce que les éléments obtenus appartiennent aux lexiques définis, en tenant compte de contraintes sur leur position dans le mot (initiale, médiane ou finale) et sur la compatibilité des éléments entre eux. Une segmentation *valide* est alors une segmentation pour laquelle l'ensemble des éléments obtenus appartient aux lexiques de la langue. À l'inverse, une segmentation *invalid*e est celle dont un (ou plus) des éléments obtenus ne se trouve pas dans ces lexiques.

Par exemple, la segmentation du mot فَدَّهَبَ /favahaba/ « puis il est parti » en فَ /fa/ (particule de coordination) + دَّهَبَ /vahaba/ (verbe ذهب) est valide, par contre la segmentation



du mot فَرَحَ /faraHa/ « il était content » en فَ /fa/ + رَحَ /raHa/ est invalide (l'élément /raHa/ n'appartient pas aux lexiques de la langue). Plusieurs analyses morphologiques valides peuvent être obtenues pour un mot donné : le mot كَتَبَ /ktb/ (sans voyelles) peut produire كَ /k/ + تَبَّ /tabba/ « comme trancher » ou كَتَبَ /kataba/ « écrire » (ce mot non voyellé présente 9 voyellations possibles [Deb98]). Dans ce cas, des connaissances de niveaux supérieurs (syntaxique, sémantique...) sont nécessaires pour lever l'ambiguïté.

Différents types de segmentations ont été proposés dans la littérature : Taibi [Tai97] propose un découpage en 3 classes : (préfixe + base + suffixe) ; Zemirli [Zem98b] en 5 classes (antéfixe + préfixe + base + suffixe + postfixe). C'est ce découpage que nous avons adopté dans le processus de segmentation (cf. chapitre 5). Ainsi, plus le nombre de classes est réduit, plus la procédure de découpage est simplifiée mais plus le nombre d'éléments dans chaque classe est élevé.

L'assignation de valeurs morphologiques aux éléments lexicaux obtenues constitue l'étape suivante. Ces valeurs ont trait au type de l'élément (verbe, adjectif, pronom, etc.), au genre (masculin, féminin), au nombre (singulier, duel, pluriel), etc. Plusieurs valeurs morphologiques peuvent être attribuées à un même élément lexical compte tenu d'un trait morphologique donné. C'est le problème de l'homographie polycatégorielle qui est traitée au niveau des modules en aval.

#### 4.2.1 Mécanisme de dérivation

##### A. La racine

Une partie importante du lexique arabe est structurée autour d'une racine : un groupement de trois consonnes (quatre dans 1 à 2 % des cas [Boh79]). Une racine, à travers le mécanisme de dérivation, peut donner naissance à une famille de mots autour d'un même *concept sémantique*. Par exemple, la racine كَتَبَ /ktb/ peut engendrer 15 mots autour du concept « écriture » : كِتَابٌ /kitAbun/ (« livre »), مَكْتَبٌ /maktabun/ (« bureau »), مَكْتُوبٌ /maktUbun/ « écrit », etc. Les lettres de la racine كَتَبَ gardent leur position dans les mots engendrés, et des voyelles (ا /A/ de كِتَابٌ /kitAbun/) ou des consonnes (م /m/ de مَكْتَبٌ /maktabun/) peuvent s'y ajouter. C'est le fait le plus caractéristique de la morphologie arabe et, plus généralement, sémitique.

Une des caractéristiques des dictionnaires de la langue arabe est que les mots y sont classés par ordre alphabétique de leur racine. La recherche du mot مَكْتَبٌ /maktabun/ « bureau » par exemple passe par trois phases : la première phase, mentale, consistera à déduire la racine كَتَبَ /ktb/ de ce mot ; la deuxième phase à trouver l'entrée كَتَبَ /ktb/ dans le dictionnaire ; la troisième à chercher le mot مَكْتَبٌ /maktabun/ dans la liste des mots dérivés à partir de cette entrée.

## B. Le schème

Le schème (الوزن) est un mot prédéfini composé de trois lettres radicales ف /f/, ع /ε/ et ل /l/, auxquelles peuvent s'ajouter des lettres additionnelles (préfixe, infixe ou suffixe). La notion de schème occupe une place importante dans le mécanisme de dérivation de l'arabe.

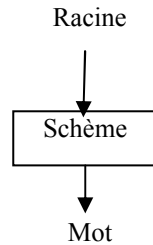


Fig. 8 : Mécanisme de dérivation en arabe.

Pour comprendre ce mécanisme, le schème peut être assimilé à un *moule* dans lequel est coulée une racine pour générer un mot (cf. figure. 8). Par exemple, le mot كَتَبَ /kataba/ est formé en coulant la racine كتب /ktb/ (sans voyelles) dans le schème فَعَلَ /faʕala/; le mot مكتوبٌ /maktUbu/ « écrit » est formé en coulant la racine كتب /ktb/ dans le schème مَفْعُولٌ /maʕʕulun/ (cf. figure 9).

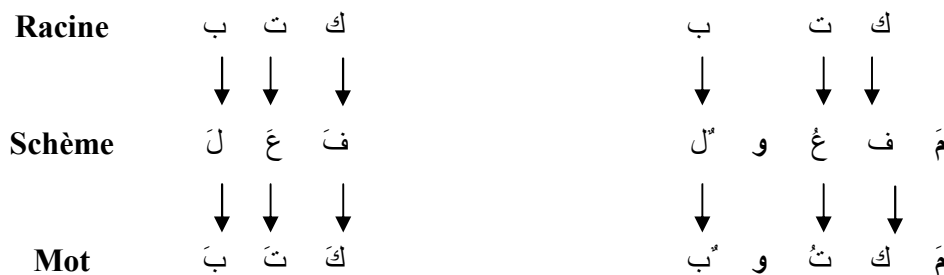


Fig. 9 : Exemple de dérivation du mot /ktb/.

La formation des mots se déroule de la manière suivante : les lettres ف /f/, ع /ε/ et ل /l/ du schème sont remplacées respectivement par les lettres ك /k/, ت /t/ et ب /b/ de la racine en gardant intactes leurs positions et la position des voyelles et des lettres additionnelles م /m/ et و /w/.

Le schème arabe intègre des informations grammaticales qui seront attribuées aux mots qu'il forme. Il confère ainsi des valeurs morphologiques, syntaxiques et quelquefois sémantiques à ces mots : en appliquant le schème فاعل /faʕil/ ayant la valeur morphologique « participe actif, masculin, singulier » à la racine كتب /ktb/, nous formons le mot significatif كاتب /kAtib/ « écrivain » possédant ces mêmes valeurs. Le schème peut être qualitatif, quantitatif avec ou sans gémation d'une de ses consonnes radicales. Il peut également être

constitué par ajout d'un préfixe, suffixe ou infixes ou encore par une combinaison de deux ou trois de ces éléments.

Les schèmes sont classés en deux catégories : les schèmes verbaux, qui permettent de former les verbes à partir d'une racine, et les schèmes nominaux, qui forment les noms. Ainsi, selon la catégorie de schème employé, il est possible de former soit un verbe ( يَكْتُبُ /yaktubu/ - (« il écrit ») à partir du schème verbal يَفْعَلُ /yafʿalu/), soit un nom ( كَاتِبٌ /kAtibun/ (« écrivain ») à partir du schème nominal فَاعِلٌ /fAʿilun/), et ceci à partir d'une même racine كَتَبَ/ktb/). Notons aussi qu'une racine ne peut pas être employée avec l'ensemble des schèmes définis dans la langue, mais uniquement avec ceux qui lui sont attestés par les linguistes. Les classes de schèmes verbaux et nominaux seront décrites plus en détail dans le chapitre 4.

#### 4.2.2. Morphologie du mot arabe

Selon la grammaire traditionnelle, le lexique arabe comprend trois catégories de mots : verbes, noms et particules (recouvrant adverbes, conjonctions et prépositions). Hormis les noms propres, les mots des deux premières catégories sont dérivés à partir d'une racine. Ils sont nommés mots réguliers ou *mots dérivables* (cf. figure 10). Les mots irréguliers, qui comprennent les mots outils et certains mots spéciaux, ne sont pas analysables en racine + schème. Ils sont appelés *mots non dérivables*.

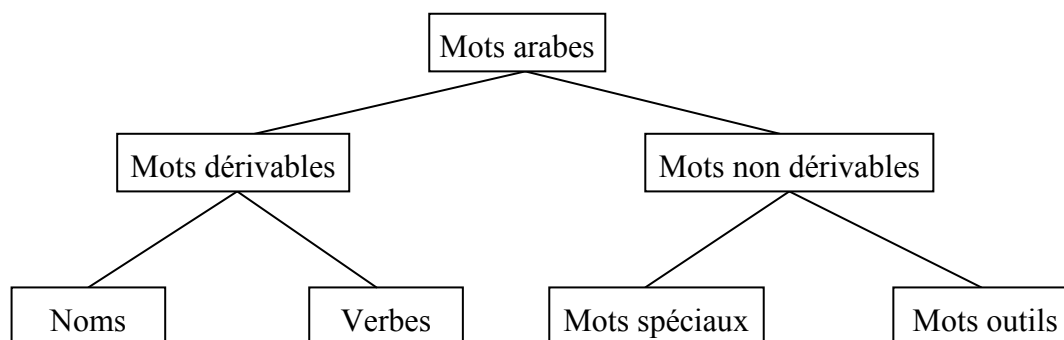


Fig. 10 : Structuration du lexique arabe.

##### A. Le nom

Le nom dans la terminologie arabe traditionnelle regroupe le substantif, l'adjectif et le pronom. Les substantifs et les adjectifs sont deux catégories qu'il est difficile de distinguer [Bla75]. Nous ne décrivons pas dans ce qui suit l'ensemble de ces classes, qui sont par ailleurs largement commentées dans la littérature [Ben93], mais quelques-unes qui aideront à comprendre notre démarche.

Un nom est dit *primitif* s'il ne dérive d'aucune racine (أَسَدٌ, /ʔasadun/ (« lion »), ثَلْجٌ /caljun/ « neige »). À l'inverse, il est dit *dérivé* s'il est formé à partir d'une racine (مَدْرَسَةٌ /madrasatun/ « école » dérive de دَرَسَ /drs/). La forme nominale peut subir deux sortes de

changement : le premier est appelé *تصريف* /taSrIf/ par les grammairiens arabes, qui peut se traduire par *changement* dans le cas des noms (passage du singulier au pluriel, etc.), et par *conjugaison* dans le cas des verbes. Le deuxième changement est appelé *إعراب* /?i?rAb/. Il concerne la *flexion casuelle* des noms (terminaison) qui est liée à leur catégorie dans la phrase. Une flexion casuelle peut être une voyelle (*fetha, damma, kasra*) ou une séquence de voyelles et de consonnes (cas du duel et du pluriel). Ces flexions casuelles, qui sont pour nous des indices très importants dans la démarche adoptée, sont recensées dans le tableau 5.

Nom	Cas sujet	Cas direct	Cas indirect
Singulier	(damma)	(fetha)	(kasra)
Duel	ان	يُن	يُن
Pluriel sain masculin	ون	يْنَ	يْنَ
Pluriel sain féminin	اتُ	اتِ	اتِ

**Tab. 5 : Les flexions casuelles de la langue arabe.**

## B. La détermination du nom

L'indétermination d'un nom arabe se traduit par le placement d'une marque de *tanwin* ( ) /an/ ( ) /un/ ( ) /in/ sur sa dernière consonne. Exemple : مُدِيرُ /mudIrun/ « directeur ». Un nom qui ne porte pas une marque de *tanwin* est par conséquent déterminé. Il peut l'être soit :

- à l'aide de l'article ال ;
- à l'aide de l'annexion d'un complément de nom : elle consiste en la juxtaposition de deux ou plusieurs termes, le premier étant déterminé par le deuxième, le deuxième par le troisième, etc.
- à l'aide d'une suffixation : un pronom se rattache au nom à la manière d'un suffixe (il correspond à l'adjectif possessif en français).

Il existe par ailleurs des noms (équivalent selon le cas à un complément circonstanciel de temps ou de lieu ou à un adverbe de temps ou de lieu) qui mettent au cas indirect le mot qui les suit.

## C. Les pronoms

Cette classe regroupe les pronoms personnels (هو, هي, هم...), les démonstratifs (هذا, هذه...) et les pronoms relatifs (الذي, الذين...). Leur nombre est stable dans la langue. Ces

pronoms s'accordent en genre et en nombre avec les noms auxquels ils se rapportent. Exemple : هذان الرجلان /havAnir rajulAni/ (le démonstratif au duel هذان s'accorde avec le nom الرجلان).

#### D. Le verbe

Le verbe arabe est soit simple, c'est-à-dire composé uniquement de lettres appartenant à sa racine ( أَكَلَ /?akala/ « il a mangé » formé à partir de la racine أَكَلَ /?kl/) en plus des voyelles, soit augmenté si une ou plusieurs lettres sont ajoutées à la racine ( مُدَرِّسٌ /mudarrisun/ formé à partir de la racine /drs/ en ajoutant la lettre م /m/). Il est le plus souvent trilitère (composé de 3 lettres).

Le verbe se conjugue à trois temps : l'accompli qui correspond aux actions passées ( نَجَحَ /najaHa/ « il a gagné »), l'inaccompli qui correspond aux actions présentes et futures ( يَنْجَحُ /yanjaHu/ « il gagne ou il gagnera ») et l'impératif ( اِنْجَحْ /?injaH/ « gagne »). Il peut être transitif, intransitif s'il n'accepte pas de complément ou transitif et intransitif à la fois selon son contexte. Enfin, il peut être soit à la voie active ( يَقُولُ /yaqUlu/ « il dit ») soit à la voie passive ( يُقَالُ /yuqAlu/ « il est dit = il se dit »), la différence résidant au niveau des voyelles.

Le verbe arabe apparaît sous des formes prédéfinies dans la langue. Ces formes, au nombre de 14, sont dérivées à partir du schème trilitère فَعَلَ /faʕala/, redoublement de la seconde lettre radicale ( فَعَّلَ /faʕʕala/), ou ajout de préfixes au début de la racine ( أَفَعَلَ /?afaʕala/) ou d'infices en son milieu ( فَاعَلَ /fAʕala/). Les schèmes dérivés ainsi obtenus expriment des sens différents de celui du schème trilitère فَعَلَ.

Selon le mécanisme de dérivation de l'arabe, l'ensemble des formes dérivées d'une racine trilitère quelconque peut être obtenu en coulant cette racine dans les moules de schèmes dérivés.

	فَعَّلَ	→	كَسَّبَ
	أَفَعَلَ	→	أَكَسَّبَ
كَسِبَ	تَفَعَّلَ	→	تَكَسَّبَ
	اِفْتَعَلَ	→	اِكْتَسَّبَ
	اِسْتَفَعَلَ	→	اِسْتَكْسَبَ

Tab. 6 : Formes dérivées à l'accompli de la racine trilitère /kasb/.

Le tableau 6 représente les formes dérivées à l'accompli de la racine trilitère كَسِبَ /kasaba/ « il possède ». Nous en tirons les remarques suivantes :

- Cinq verbes à l'accompli peuvent être formés à partir de la racine /kasaba/ en coulant celle-ci dans les schèmes *فَعَّلَ*, *أَفْعَلَ*, *تَفَعَّلَ*, *اِفْتَعَلَ* et *اِسْتَفَعَلَ*. Les correspondances entre cette même racine et les autres schèmes dérivés définis dans la langue (14-5 = 9 schèmes) ne sont pas attestées par les linguistes ;
- La formation de ces verbes se fait en remplaçant respectivement les lettres *ل*, *ع* et *ف* des schèmes en question par les lettres *ب*, *س* et *ك*, en gardant intact la position des voyelles et des lettres additionnelles ;
- Les verbes formés à partir de la racine /kasaba/ ont des sens différents.

Les verbes dont la racine renferme une des trois lettres *أ* /ʔ/, *و* /w/ et *ي* /y/ sont appelés verbes *faibles* ou *malades*. Dans le processus de dérivation, ces lettres dites *faibles* sont modifiées ou supprimées.

Racine		ذ	خ		أ	
Schème		لَ	عَ	تَ	فَ	اِ
Mot		ذ	خ	تَ	أ	اِ
Verbe		ذ	خ	تَ	ت	اِ

↓  
Dérivation

Tab. 7 : Exemple de génération du verbe /ʔittaxava/.

Le tableau 7 représente le processus de formation du mot *اِتَّخَذَ* à partir de la racine trilitère *أخذ* /ʔxd/. Cette racine est coulée dans le schème *اقتعل* et donne le mot *اِتَّخَذَ*. La lettre faible /ʔ/ de cette racine est ensuite remplacée par la lettre *ت* /t/ selon des règles phonologiques bien établies. Les différents types de verbes malades sont présentés dans des ouvrages classiques de la littérature [Bla75] [Ben93]. Par opposition, les verbes qui ne renferment pas de lettres faibles sont appelés verbes *forts* ou verbes *sains* (*كسب* /kasaba/ est un verbe sain).

À l'instar des noms, la déclinaison des verbes peut renseigner sur leur mode, leur temps et leur nombre :

- la *damma* termine le verbe singulier à l'inaccompli et au futur. Ex : *يَذْهَبُ* /yavhabu/ (« il part ») ;
- la *fatha* termine le verbe à l'accompli (*خَرَجَ* /xaraja/ « il est sorti ») et au subjonctif (*أَنْ يَخْرُجَ* /ʔan yaxruja/ « il faut qu'il sorte ») ;
- la *soukoun* termine le verbe apocopé (*لَمْ يَذْهَبْ* /lam yavhab/ « il n'est pas parti »).

Les verbes au subjonctif et les verbes apocopés sont précédés par des particules qui commandent leurs différentes terminaisons.

## E. Les particules

Les particules sont des mots invariables qui indiquent les articulations de la phrase. Ils sont en nombre limité et peuvent accompagner les noms, les verbes ou les deux à la fois. Ces particules peuvent être classées dans les catégories suivantes :

- Particules employées devant un nom qui régissent le cas indirect (prépositions) ;
- Particules de coordination ;
- Particules de négation ;
- Particules interrogatives ;
- Particules de condition ;
- Particules de subjonctif ;
- Particules de l'apocopé ;

### **4.3. La grammaire en tronçons : application à l'arabe**

Comme nous l'avons évoqué dans la première partie, l'analyse syntaxique doit être guidée par les contraintes de la SAT. C'est dans ce contexte qu'une grammaire en tronçons est proposée : fondée sur une analyse superficielle et non-exhaustive du texte, cette grammaire consiste à diviser la phrase en groupes de mots non récursifs, baptisés *chunks* en anglais [Abn91], *tronçons* en français [Bou97], sans nécessairement les mettre en relation les uns avec les autres. Les mots appartenant à un même tronçon se caractérisent par des liens syntaxiques forts : ainsi, leur ordre dans le tronçon est rigide comparé à l'ordre des tronçons dans la phrase, qui est relativement flexible.

D'un point de vue prosodique, le tronçon ne peut être scindé ni par une pause ni par une frontière intonative : il trouve son équivalent à l'oral sous la forme de groupes accentuels, constitués d'un mot *accentogène* et de mots clitiques (sans accent lexical) périphériques [Dej98]. Par ailleurs, la pertinence de cette unité dans la hiérarchie mot-tronçon-phrase a été démontrée dans différentes langues [Gig98]. La question que nous nous sommes dès lors posée est la suivante : comment délimiter ces tronçons en arabe ? Notre investigation a porté sur l'étude des procédés grammaticaux par lesquels les mots arabes sont rattachés les uns aux autres.

L'arabe peut être caractérisé par trois faits syntaxiques [Boh79] :

#### **La proéminence du verbe**

Il existe deux types de phrase en arabe : la phrase nominale qui commence par un nom et la phrase verbale par un verbe, les deux phrases pouvant être introduites par une particule. Le verbe occupe une place de choix dans une phrase verbale (canoniquement *verbe* + *sujet* +

*complément direct + complément circonstanciel*) dans la mesure où il organise la place de ces constituants. Sous l'influence des langues européennes, l'ordre des constituants dans la phrase peut changer (il devient sujet + verbe + complément direct + complément circonstanciel).

### L'ordre des unités

Certaines unités comme les couples *nom + complément du nom* et *nom + épithète* se combinent selon un ordre rigide. Il existe par ailleurs des unités à régime fixe, c'est-à-dire des mots exigeant à la suite une classe ou une flexion précise (*préposition + complément indirect*, *particule de négation + verbe*), sur lesquelles nous nous sommes beaucoup appuyés pour la désambiguïisation contextuelle.

### L'accord entre les unités

Ces accords ont trait à la variation du genre (masculin ou féminin) et du nombre (singulier, duel, pluriel). Le verbe s'accorde avec le nom ou le pronom suivant d'une part, et l'adjectif ou l'épithète avec le nom ou le pronom auxquels ils sont rattachés d'autre part.

La définition du tronçon en arabe découle directement de ces trois faits syntaxiques : toute séquence de mots constituée d'un verbe ou de noms, obéissant à un ordre rigide et à des contraintes d'accords fortes, est assimilée à un tronçon. À partir de là, nous avons défini quatre types de tronçons (*cf.* figure 11) :

- Tronçon verbal

Ce type de tronçon regroupe le verbe et d'éventuelles particules le précédant. L'élément central de ce tronçon est le verbe.

- Tronçon sujet

Ce tronçon regroupe les formes *nom sujet + complément du nom* et *nom sujet + épithète*. L'élément central de ce tronçon est le nom, qui peut être précédé par une particule.

- Tronçon objet

Ce tronçon regroupe les formes *nom objet + complément du nom* et *nom objet + épithète*. L'élément central dans ce tronçon est le nom, qui peut être précédé par une particule.

- Tronçon indirect

Ce type de tronçons regroupe les formes *préposition + complément indirect*, la tête restant nominale.

( ذَهَبَ )<sub>1</sub> ( الْوَلَدُ )<sub>2</sub> ( إِلَى الْمَدْرَسَةِ )<sub>4</sub> ( بَعْدَ الْأَكْلِ )<sub>4</sub>  
 ( الْوَلَدُ الصَّغِيرُ )<sub>2</sub> ( مَسْرُورٌ )<sub>2</sub>

Fig. 11 : Exemple de découpage en tronçons (entre parenthèses).



#### **4.4. L'analyse morpho-syntaxique**

Rappelons qu'une analyse morpho-syntaxique introduit des contraintes contextuelles sur les valeurs morphologiques attribuées aux éléments lexicaux, ce qui réduit les ambiguïtés au niveau lexical et structurel. L'analyse morpho-syntaxique que nous avons développée reprend les principes de Vergne [Ver99]. L'approche repose sur l'étiquetage par défaut, la propagation de déductions contextuelles et l'utilisation d'un lexique partiel. Elle a déjà été implémentée dans un système de SAT pour répondre notamment aux contraintes de ces systèmes en terme de rapidité d'exécution [Van99a].

- Étiquetage par défaut

Remplacer l'analyseur morpho-lexical par un tagger ne résout que partiellement le problème de la combinatoire. Au mieux, une seule étiquette est retenue. L'étiquetage par défaut est une technique pour supprimer la combinatoire lexicale en ne gardant qu'une seule étiquette par mot : si un mot est ambigu, une étiquette par défaut lui est attribuée. Le choix de cette étiquette est établi à partir d'observations de corpus : « Nous faisons l'hypothèse que la valeur par défaut (venant du lexique), jusqu'à preuve du contraire (apportée par le contexte), est une propriété générale des langues, permettant une économie dans les processus de communication langagiers » [Ver98a, page 3]. Les règles de déductions interviennent en aval pour confirmer la valeur attribuée par défaut, ou au contraire, modifier cette valeur en fonction du contexte d'apparition du mot.

- Ressources partielles

Comme nous l'avons évoqué, l'analyse syntaxique traditionnelle utilise des lexiques supposés exhaustifs. Cette supposition sous-entend que l'ensemble des mots à analyser soit connu et répertorié. Dans la pratique, il en est tout autrement : il existe dans les documents une importante masse de mots inconnus (mots d'emprunt, techniques, erronés, etc.) [Gig98].

Une approche guidée par le lexique n'est guère appropriée dans le cas d'une application à large couverture, à cause des erreurs engendrées par les mots inconnus. La stratégie adoptée doit tenir compte de ces exceptions en explorant d'autres sources de connaissances plus stables, comme le contexte immédiat. Il est cependant incontestable que le lexique peut aider à la résolution de problèmes là où une approche non-lexicaliste échoue. Un lexique de mots outils serait même incontournable [Bou97].

- Déduction contextuelle

Les déductions contextuelles représentent une ressource supplémentaire, en synergie avec l'incomplétude des lexiques. Vergne [Ver99] délimite d'abord les tronçons sur la base de mots outils et applique ensuite des règles contextuelles à l'intérieur de ces tronçons. Comme

les positions des mots à l'intérieur du tronçon sont rigides, les règles appliquées utilisent des critères positionnels pour confirmer ou infirmer les étiquettes établies. Elles peuvent s'utiliser de manière *négative*, en supprimant des étiquettes ne pouvant pas apparaître dans certains contextes (la particule *أَنْ* /ʔan/ n'est jamais suivie par un nom), ou de manière *affirmative*, en introduisant des étiquettes dans des contextes sûrs (la préposition *فِي* /fi/ est toujours suivie d'un nom au cas indirect). Il s'agit dans le premier cas de *déductions négatives*, et dans le deuxième de *déductions affirmatives*.

#### **4.5. Travaux dans le domaine**

Nous présentons ci-dessous quelques travaux sur la morphologie en arabe.

- Les travaux de SAROH [Sar90]

SAROH a développé une base de données lexicale où chaque entrée contient une racine qui pointe vers des règles de dérivation et de flexion. Les racines au nombre de 2000 permettent par ce système de règles de générer 200.000 formes fléchies. Des informations morphologiques sur le type (nominale ou verbale), la catégorie (saine, malade), etc. sont associées à chaque racine. Cette base lexicale se situe au centre d'un système modulaire dont les tâches sont les suivantes :

- Un module d'analyse morphologique pour la segmentation des mots et l'extraction de leur racine ;
- Un module de génération de mots ;
- Un module d'analyse syntaxique dont la grammaire est formalisée sous forme d'un système expert ;
- Un module de conjugaison des verbes.

Une des applications de ce système est la TOP de textes arabes voyellés.

- Les travaux de Zemirli [Zem00]

Dans le cadre du projet MEDIA, à l'Institut National d'Informatique d'Alger, un système d'analyse et de génération automatique de la langue arabe a été développé. L'architecture de la base lexicale repose sur le même principe que celle de SAROH : racine + règle. Le plus de ce système est l'implémentation de règles phonologiques pour rendre compte des transformations liées aux lettres faibles (ajout, suppression ou modification) dans les processus d'analyse et de dérivation.

- Les travaux du Centre de Recherche Scientifique et Technique pour le Développement de la Langue Arabe d'Alger (C.R.S.T.D.L.A)

Les travaux du C.R.S.T.D.L.A ont pour objectif le développement d'une boîte à outils logicielle pour le traitement automatique de la langue. Tous ces travaux reposent sur le

modèle linguistique basé sur la théorie néo-khalilienne [Had79]. Parmi eux, les travaux de TAIBI qui a développé un système d'analyse morpho-syntaxique en vue de la correction d'erreurs orthographiques [Tai97].

Citons également les travaux de Debili [Deb98]. Tous ces systèmes ont le même objectif, à savoir le développement d'un système d'analyse et de génération pour la langue arabe. Pour cela, ils utilisent une base de données lexicales supposée exhaustive selon la représentation racine + règle le plus souvent. En revanche, ils diffèrent quant à la technique de segmentation utilisée.

## CHAPITRE 5 : Méthode d'analyse

Comme nous l'avons vu au chapitre 4, l'analyse morpho-syntaxique que nous avons développée reprend les principes de Vergne [Ver99]. Comme le montre la figure 12, la première étape de cette analyse est l'étiquetage morpho-syntaxique qui va assigner à chaque *token* une étiquette unique, en utilisant une base de ressources lexicales partielles couplée avec une base de règles de déduction contextuelle. La séquence de tokens étiquetée est ensuite passée au module suivant de parenthésage syntaxique qui a pour tâche de découper cette séquence en tronçons.

Parallèlement à l'analyse morpho-syntaxique, le texte orthographique est converti en une chaîne de phonèmes par le module de TOP qui fera l'objet du chapitre suivant. La chaîne phonétique ainsi que la séquence de tronçons se rejoignent au niveau de l'interface syntaxe-prosodie dont les fonctions sont la distribution des pauses et le calcul des paramètres prosodiques, à savoir, la fréquence du fondamental et la durée phonémique.

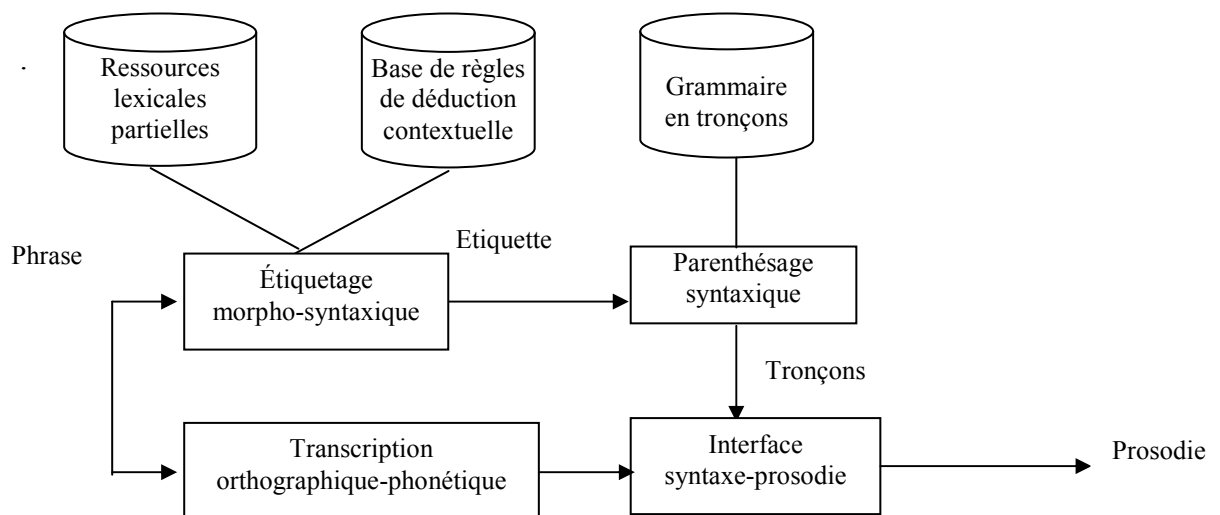


Fig. 12 : Diagramme bloc de l'analyse linguistique.

### 5.1. L'étiquetage morpho-syntaxique

L'étiquetage *morpho-syntaxique* a pour but d'associer une *étiquette grammaticale* à chaque mot de la phrase. La première étape est alors de définir un jeu d'étiquettes, qui soit adapté au découpage en tronçons arabes, tels que nous les avons définis. En effet, l'étiqueteur morpho-syntaxique doit fournir au module de découpage en tronçons toutes les informations grammaticales dont il a besoin pour mener à bien son processus. La question que nous nous sommes alors posée est la suivante : quelles sont les informations grammaticales utiles au découpage de la phrase en tronçons arabes ? La réponse à cette question est fortement liée au choix des étiquettes utilisées.

## A. Choix des étiquettes

Tout d'abord, il faut rappeler que nous travaillons dans le cadre de textes voyellés. La flexion casuelle des mots constitue un élément important dans la tradition grammaticale de l'arabe car, comme nous l'avons mentionné, elle renseigne sur la fonction des noms, le mode des verbes (passif, actif), etc. Par ailleurs, cette langue dispose d'un ensemble d'étiquettes morphologiques (participe actif, participe passif, nom verbal...), où la notion de *schème* occupe une place importante, sans référence à la position des mots dans la phrase [Deb98].

C'est dans ce contexte que nous avons défini une liste d'étiquettes morphologiques, au nombre de 23, qui rendent compte de la nature du mot (verbe, nom, particule) ainsi que, pour les noms, de leur flexion casuelle (cas sujet, objet ou indirect), de leur état déterminé/indéterminé et du type de détermination (par l'article, par suffixation d'un pronom personnel ou par annexion d'un complément du nom). Ce choix est étroitement lié au regroupement en tronçons. Ainsi, tout un éventail de traits morphologiques n'est pas nécessaire : il n'est pas utile par exemple de connaître l'aspect (accompli, inaccompli) des verbes ou le genre (masculin, féminin) des noms dans le processus de découpage.

La notation des parties du discours retenue ici est inspirée du projet européen MULTTEXT [Cam98], même si, comme l'évoque Blachère, faire correspondre les catégories indo-européennes dans le cadre de l'arabe n'est pas toujours aisé [Bla75]. Observons les différents constituants des tronçons un par un, et dressons les informations nécessaires à leur calcul :

- Tronçon verbal

Etiquettes	Description	Exemple
V	Verbe (personnel simple)	ذَهَبَ ، ذَهَبْتَ ، سَيَذْهَبُ
Vp	Verbe avec préfixe (conjonction de coordination)	فَذَهَبَ
Vs	Verbe avec suffixe(s) (pronom complément)	كَرَّمَكُ ، كَرَّمَهُمُ
Vps	Verbe avec préfixe (conjonction de coordination) et suffixe(s) (pronom complément)	فَكَرَّمَهُمُ

**Tab. 8 : Liste des étiquettes verbales.**

Le tronçon verbal, rappelons-le, recouvre les formes *verbe* et *particule + verbe*. Comme la flexion casuelle d'un verbe ne détermine pas formellement sa catégorie, la prise en compte de cette information n'est pas pertinente pour l'étiquetage du verbe. Ici, la détection du verbe est uniquement basée sur la notion de schème, comme nous le détaillerons ultérieurement. De plus, nous avons introduit des étiquettes verbales qui rendent compte des préfixes et suffixes attachés au verbe (cf. tableau 8), exploitées par ailleurs pour l'assignation de l'accent lexical et des pauses.

- Tronçon nominal

Étiquettes	Description	Exemple
Nsi	Nom sujet indéterminé (تنوين)	مَكْتَبٌ ، طِفْلٌ
Nsd	Nom sujet déterminé par l'article (ال)	الْوَالِدُ ، الطَّائِرُونَ
Nsa	Nom sujet déterminé par annexion (complément du nom)	وَالِدٌ ، سَائِقٌ
Nss	Nom sujet déterminé par suffixation (pronom personnel)	حَجْمَةٌ ، مَقْرُهَا
Noi	Nom objet indéterminé	كَنْزًا ، حَسَنَةً
Nod	Nom objet déterminé par l'article (ال)	الْوَالِدَ ، الطَّائِرِينَ
Noa	Nom objet déterminé par annexion (complément du nom)	مَصْلَحَةً ، مَدْخَلٌ
Nos	Nom objet déterminé par suffixation (pronom personnel)	وَقْتَهُ ، مَكْتَبَهَا
Nii	Nom indirect indéterminé	شَهْرٌ ، حَلَوِيَّاتٍ
Nid	Nom indirect déterminé par l'article (ال)	الْمَدْرَسَةَ ، الْمَنْزِلَ
Nia	Nom indirect déterminé par annexion (complément du nom)	دَاخِلٌ ، جَنَاحٍ
Nis	Nom indirect déterminé par suffixation (pronom personnel)	طَرِيقَهُ

**Tab. 9 : Liste des étiquettes nominales.**

Le tronçon nominal recouvre les formes *nom + épithète* et *nom + complément de nom*. Les informations sur la flexion casuelle et le trait *déterminé/non déterminé* du nom, de l'épithète et du complément de nom permettent de construire ces deux formes de tronçons. Rappelons que le complément de nom peut se manifester soit par annexion d'un nom déterminé au cas indirect, soit par suffixation d'un pronom personnel. Nous avons choisi de classer les étiquettes qui expriment ces informations morphologiques en trois sous-classes, selon les flexions au *cas sujet*, au *cas objet* et au *cas indirect*. La liste de ces étiquettes est présentée dans le tableau 9.

- Tronçon indirect

Le tronçon indirect recouvre la forme *préposition + nom au cas indirect*, celui-ci pouvant être déterminé ou indéterminé. Les prépositions sont répertoriées dans un lexique de formes (forme canonique + forme fléchie) qui est consulté pour leur étiquetage. La détection du nom est pour sa part basée sur sa flexion casuelle et son caractère déterminé/non déterminé. Rappelons qu'une préposition peut être attachée au nom auquel elle se rapporte *لِلوَالِدِ* /lilwaladi/ (« pour le garçon ») et que celui-ci peut se lier par suffixation à la préposition *فِيهِ* /fih/ (« en son sein »). Les étiquettes définies, présentées dans le tableau 10, rendent compte de toutes ces situations.

Étiquettes	Description	Exemple
Si	Préposition gouvernant le cas indirect (حروف الجر)	فِي ، عَلَى ، إِلَى ، عَنْ ، مِنْ ، أَمَامَ ، فَوْقَ ، تَحْتَ ، حِينَ ، مُدَّةً ، مُنْذُ ، مُذْ ، بَعْدَ ، قَبْلَ ، حَوْلَ ، وَرَاءَ ، بَيْنَ ، عِنْدَ
Sii	Préposition + nom indirect indéterminé	بِدَارِ ، كَمَنْزِلِ
Sid	Préposition + nom indirect déterminé par l'article (ال)	لِلوَالِدِ ، بِالْمَنْزِلِ
Sia	Préposition + nom indirect déterminé par annexion (complément du nom)	لِمَنْزِلِ
Sis	Préposition + nom indirect déterminé par suffixation (pronom personnel) ou préposition + suffixe(s)	فِيهِ ، لِوَالِدِهَا ، فَوْقَهُ ، عِنْدَهَا

Tab. 10 : Liste des étiquettes au cas indirect.

Il nous reste à définir des étiquettes correspondant aux particules de la langue arabe. L'étiquette **P** recouvre l'ensemble des mots outils non référencés par les autres étiquettes : les particules interrogatives, les particules affirmatives, les particules de négation, les particules d'insistance, les particules de coordination, les démonstratifs, les pronoms personnels isolés, les pronoms relatifs ainsi que toutes les autres particules, hormis les prépositions. De plus, nous avons volontairement inclus les pronoms dans la classe des particules.

Outre les étiquettes définies, des étiquettes temporaires (internes, n'apparaissant pas en surface) sont introduites, par exemple Noa pour les pluriels en *ات* /Ati/ ou les duels en *ين* /yn/ dont seul le contexte permet de trancher entre objet direct et indirect — sinon, c'est Noa qui est assigné par défaut. Des étiquettes mixtes sont également définies pour des cas où il est difficile de déterminer, par exemple, si un mot est sujet ou objet, en veillant à ne pas le faire

précéder par une frontière de tronçon : pensons aux mots terminés en *ى* ou en *ي* (qui peuvent être un pronom lié de la 1<sup>ère</sup> personne du singulier) ou aux noms propres non signés. Certains schèmes qui peuvent être identifiés comme nominaux ou verbaux reçoivent un traitement analogue.

## B. Ressources lexicales

La langue arabe souffre d'un manque en ressources électroniques, ce qui est un handicap non négligeable pour le traitement automatique. La mise en œuvre d'une base lexicale est une tâche de longue haleine qui dépasserait le cadre de notre travail. Nous utilisons dans ce travail (cf. tableau 10) des lexiques partiels de mots grammaticaux (particules), de formes verbales, de déclinaisons nominales et de mots spécifiques pour le traitement des exceptions. Il faut préciser que nous n'utilisons pas de lexique de racines arabes.

Préfixes	Suffixes	Schémes verbaux (sains)	Schémes verbaux (malades)	Déclinaisons nominales	Particules	Mots spécifiques
11	26	14	19	15	157	60

**Tab.10 : Ressources utilisées pour l'analyse morphologique.**

### ▪ Lexique des formes verbales

Le schème dans la langue arabe peut être considéré comme un *modèle* qui spécifie la morphologie des mots de cette langue. La détection du verbe se fait en comparant sa forme aux différents schèmes verbaux existants. Cette approche suscite cependant quelques réflexions :

- Tous les verbes arabes, en particulier les verbes malades, ne sont pas directement comparables aux schèmes verbaux définis dans la langue. En effet, ces derniers subissent des transformations dans leur dérivation, qui se traduisent par la suppression/modification d'une ou de plusieurs de leurs lettres.
- Certains mots peuvent revêtir une forme verbale, sans être pour autant des verbes (nom de couleur, nom propre...). Exemple : أَحْمَرُ /?ahmaru/ (« rouge ») a la forme أَفْعَلُ (verbe à l'inaccompli à la première personne du singulier).
- Cette approche qui se base simplement sur la forme des verbes n'a aucun moyen de vérifier leur appartenance à la langue. Par exemple, le mot يَلْتَصُّ /yaltaSu/ a bien une forme verbale (يَفْعَلُ), mais n'appartient pas à la langue.

Nous allons tenter de justifier notre approche de détection du verbe, en présentant les solutions adoptées face aux problèmes soulignés ci-dessus. À côté des 14 schèmes verbaux relatifs aux verbes sains, appelés *schèmes profonds*, nous avons défini des schèmes verbaux



supplémentaires (19) qui recouvrent les verbes malades. Ces derniers sont appelés *schèmes de surface* [Tai97]. Cette première solution nous permet donc de comparer directement les verbes malades aux schèmes profonds définis. Zemirli [Zem98b] a quant à lui défini des règles phonologiques pour transformer les verbes malades en une forme qui soit comparable aux schèmes profonds.

Pour le second point énoncé, qui concerne les noms dont la forme est celle d'un verbe, nous en avons introduit un certain nombre dans un lexique de mots spécifiques qui est présenté plus loin dans cette section. Pour le dernier point, à savoir la non-vérification de l'appartenance des verbes à la langue, comme l'approche définie ici s'inscrit dans le cadre d'une analyse morpho-syntaxique pour la synthèse vocale, le système doit de ce fait traiter et prononcer tous les mots qui se présentent à lui, même ceux n'appartenant pas à la langue. D'une manière générale, si un mot a une forme verbale et s'il n'est pas répertorié dans le lexique des mots spécifiques, il reçoit par défaut l'étiquette verbale. Les règles contextuelles examineront ensuite cet étiquetage par défaut.

Le lexique de schèmes verbaux recouvre toutes les formes verbales arabes saines et malades. Ces schèmes rendent compte des différents temps du verbe (accompli, inaccompli, impératif), de ses modes (passif, actif) ainsi que de tous les augments (proclitique et enclitique) susceptibles de se rattacher à lui. Les formes verbales ont la structure suivante :

{Antéfixe} {Préfixe} {Base} {Suffixe} {Postfixe}.

- Les antéfixes se lient au début du mot. Ils sont au nombre de sept : ل, س, ف, ب, ك, أ, ال. Certains de ces antéfixes se combinent exclusivement avec le verbe, d'autres exclusivement avec le nom et d'autres encore à la fois avec le verbe et le nom. Leur présence renseigne ainsi sur la nature de la base (verbe/nom) et pour le verbe sur son mode et son temps. Les antéfixes peuvent se combiner enfin entre eux pour former des antéfixes composés.
- Les préfixes, au nombre de quatre (أ, ن, ت, ي) se lient avant la base et renseignent sur les modes inaccompli et impératif du verbe. Ils peuvent se combiner entre eux ou avec les antéfixes.
- Les suffixes se lient après la base. Ils sont au nombre de seize : نّ, ات, و, ي, تا, ين, أ, وا, ت, ن, تما, تم, تن, نا, ان, ون, . Les suffixes ont pour rôle de fournir la terminaison du verbe conjugué à tous les temps. Ils peuvent se combiner entre eux pour former un suffixe composé.
- Les postfixes se lient en fin de mot. Ils sont au nombre de dix : ه, ها, هما, هم, هنّ, ك, ني, كما, كم, كنّ. Ils s'accordent aux verbes pour constituer des pronoms directs, ou se raccorder aux noms pour constituer des compléments de nom. Ils peuvent se combiner pour former un postfixe composé.

La base correspond au mot qui est obtenu en coulant une racine verbale dans un schème prédéfini. Ainsi, la base du mot سيسألونها de l'exemple suivant est dérivée de la racine سأل/s?l/.

Exemple :

Postfixe	Suffixe	Base	Préfixe	Antéfixe
ها	ون	سأل	ي	س

Mis à part la base, les autres composants des formes verbales peuvent s'alterner pour rendre compte de tous les agencements possibles. Certaines combinaisons d'éléments sont interdites. Par exemple, le préfixe ت ne se construit jamais avec le suffixe تما. A cet effet, nous avons défini à l'instar de Zemirli [Zem98b] des tables de compatibilité entre ces éléments qui vont servir de base à la construction des formes verbales.

	أ	ن	ت	ي
أ	0	0	0	0
ال	0	0	0	0
ل	1	1	1	1
س	1	1	1	1
ف	1	1	1	1
ب	0	0	0	0
ك	0	0	0	0

Tab. 11 : Table de compatibilité antéfixe-préfixe.

Le tableau 11 représente les compatibilités antéfixe-préfixe, où la valeur 0 signifie que les deux éléments sont incompatibles et ne peuvent donc pas composer un mot ensemble.

Exemple de formes verbales

Les lettres ف, ع, ل sont des lettres génériques qui représentent l'ensemble des consonnes de l'alphabet arabe. Avec le schème dérivé فَعَل, nous pouvons obtenir les formes suivantes :

Schème verbal accompli

فَعَل → فَعَّلَ (antéfixe ف + schème)  
 فَعَل → فَعَّلْتَ (schème + suffixe ت)  
 فَعَل → فَعَّلَكَ (antéfixe ف + schème + postfixe ك)

Schème verbal inaccompli

يُفَعَل → يُفَعِّلُ (antéfixe ف + préfixe ي + schème)  
 يُفَعَل → يُفَعِّلَنَّ (préfixe ي + schème + suffixe ن)  
 يُفَعَل → يُفَعِّلَكَ (préfixe ي + schème + postfixe ك)

Schème verbal impératif

فَعَل → فَعِّلْ (antéfixe ف + schème)  
 فَعَل → فَعِّلُوا (schème + suffixe وا)  
 فَعَل → فَعِّلْهُ (schème + postfixe ه)

- Lexique de formes nominales

Les formes nominales rendent compte de l'agglutination à droite des prépositions aux noms, du caractère déterminé/non déterminé des noms et de leurs déclinaisons. Elles ont la structure suivante :

{préposition} {article} {nom} {suffixe} {postfixe}

Les éléments de ces différentes classes sont : { préposition }={ك ، ب ، ل ، ال} ,  
{article}={ال} , {suffixe}={ن ، ات ، و ، ي ، تا ، ي ، و ، ات ، ن} ,  
{postfixe}={ك ، ني ، كما ، كم ، كن ، ه ، ها ، هما ، هم ، هن} .

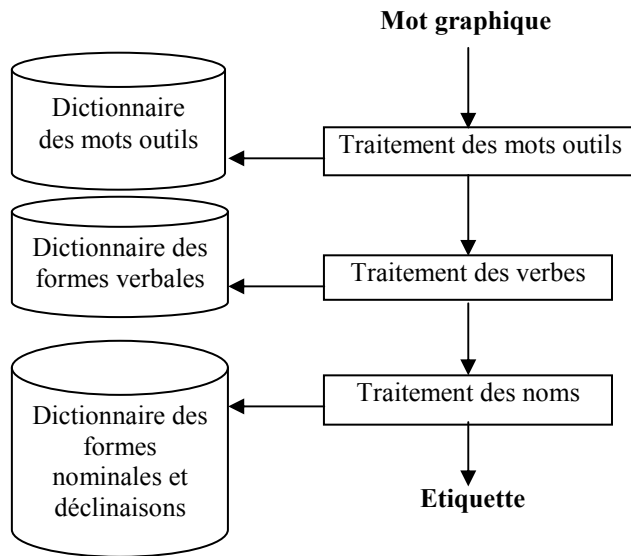
- Lexique de mots grammaticaux et de mots spécifiques

Ce lexique contient les mots outils dans leur forme canonique et fléchie (qui forment un ensemble très stable, même s'ils peuvent s'amalgamer avec des affixes pour donner naissance à de nouveaux mots) ainsi que des mots spécifiques qui aident l'analyse (mots terminés en هـ, par exemple, pour éviter la confusion avec le pronom personnel, noms masculins de couleur, certains noms propres, etc.).

### C. Analyse morphologique

L'analyse morphologique se déroule dans l'ordre suivant : elle commence par consulter le lexique des mots grammaticaux et des mots spécifiques, puis les formes verbales si la première phase échoue et enfin les formes nominales en cas d'échec des deux premières phases (cf. figure 13). Ainsi, à chaque niveau d'analyse, le mot est comparé aux différentes *entrées* du lexique, puis, en cas de succès, celui-ci reçoit l'étiquette grammaticale correspondante. Dans le cas contraire, il est transmis aux niveaux supérieurs.

Contrairement aux mots grammaticaux, le traitement des verbes et des noms n'est pas déterministe : une entrée verbale ou nominale peut désigner deux ou plusieurs étiquettes. Par exemple : la forme verbale أَفْعَلُ peut correspondre à un nom (أَكْبَرُ) ou à un verbe (الْعَبُّ) ; la flexion casuelle يَنْ (au duel) peut correspondre à un nom au cas direct ou un nom au cas indirect. Aussi, pour déterminer la terminaison des mots دَارِهِ /dArihi/ (« sa maison ») et اللَّهُ /?allahu/ (« Dieu »), il est important de savoir si la lettre هـ /h/ est un postfixe (pronom personnel) ou si elle fait partie du mot ; si la lettre كْ /k/ de كَبِيرٍ /kabIrin/ (« grand ») et /kadAri/ est une préposition attachée ou si c'est une lettre radicale.



**Fig. 13 : Analyse morphologique.**

Pour cette raison, les mots ambigus ont été recensés dans un lexique, mais ceci ne résout que partiellement le problème en raison de leur nombre important dans la langue. Pour pallier ce problème, nous leur avons attribué une seule étiquette par défaut de la manière suivante :

- Les mots qui ont une forme verbale reçoivent une étiquette verbale (si, bien sûr, ils sont absents du lexique des mots spécifiques) ;
- Les noms qui se terminent par une flexion ambiguë au cas indirect reçoivent une étiquette au cas direct ;
- Les noms dont les éléments (en début et en fin) s'apparentent à des affixes sont considérés comme des noms composés (préposition + nom et nom + pronom personnel).

*Algorithme d'analyse morphologique*

```

DEBUT
LIRE (mot)
Procédure_Tag_Mot_outils (mot, étiq.)
SI etiq = vide
  Procédure_Tag_Verbe (mot, étiq.)
  SI etiq = vide
    Procédure_Tag_nom (mot, étiq.)
  FINSI
FINSI
Renvoyer_Etiquette (mot, étiq.) ;
FIN
  
```

## 5.2. Désambiguïsation

Les règles de déduction contextuelle ou de déduction locale, interviennent après le *tagging* et ont pour tâche de vérifier l'étiquetage établi, compte tenu du contexte d'apparition des mots. Comme nous l'avons évoqué précédemment, l'ordre des tronçons à l'intérieur d'une phrase obéit à des contraintes relativement souples, comparé à celles agissant sur les mots à l'intérieur de ces groupes. De ce fait, les seules déductions fiables sont celles qui n'utilisent que les contraintes locales fortes [Gig98] dont le contexte est restreint à un seul tronçon. Ceci suppose que les frontières de tronçons soient connues au préalable.

L'analyseur de Vergne délimite les tronçons avant l'application des règles de déductions sur la base de mots outils. Dans notre cas, cette segmentation n'est pas évidente pour l'ensemble des tronçons, hormis les tronçons indirects (introduits par une préposition) et verbaux (si ceux-ci sont introduits par une particule invitant un verbe). Nous avons néanmoins émis des contraintes sur l'utilisation des règles pour s'accorder avec le principe de fiabilité des déductions dans les contextes locaux.

Les règles de déduction contextuelle proposées ici sont de type affirmatif : dans tel contexte, tel mot a telle étiquette. Ces règles sont locales — elles agissent sur un mot et ses proches voisins : leur portée est de 2 à 3 mots maximum. Le contexte examiné par ces règles n'excède jamais celui du tronçon, bien que nous ne connaissions pas encore ces limites : il est possible de savoir que deux éléments appartiennent au même tronçon, sans connaître les limites de celui-ci. Elles ne s'appliquent pas entre les mots susceptibles d'appartenir à des tronçons différents.

L'observation d'un corpus d'apprentissage a permis l'extraction de ces règles contextuelles, lesquelles seront appliquées sur des corpus réels pour la validation de nos résultats d'étiquetage. Ce corpus est issu de MULTTEXT, qui a été adapté à l'arabe et voyellé par un expert : il en résulte 40 passages de 8-10 phrases reliées par une structure thématique cohérente, comptant au total 2598 mots.

Les déductions contextuelles sont exprimées dans le même formalisme que l'analyse morphologique, à travers une vingtaine de règles de réécriture. Elles sont écrites pour *flex* – langage de traitement de chaînes de caractères dont nous avons déjà parlé, qui permet une maintenance facile de la base de règles (cf. chapitre 6). Ceci est particulièrement important, car les règles doivent être ordonnées : par exemple, la règle qui réécrit Nod (nom objet déterminé) en Nid (nom indirect déterminé) doit intervenir avant les règles appelant comme contexte une étiquette Nid.

## 5.3. Parenthésage syntaxique

Le problème est de savoir quelles sont les séquences d'étiquettes susceptibles d'appartenir à un même tronçon. Nous avons défini une relation de compatibilité (si deux

étiquettes successives sont compatibles, alors elles appartiennent au même tronçon), qui est exprimée dans une matrice dont chaque ligne (resp. chaque colonne) renvoie à l'étiquette du mot courant (resp. à l'étiquette du mot suivant). Les étiquettes sont réparties en sept classes, correspondant aux cas sujet, objet, indirect, aux verbes et aux particules de type S, P et C.

	Nsi	Nsd	Nsa	Nid	Nii	Nis	X
Nsi	0	1	1	1	1	1	1
Nsd	1	0	1	1	1	1	1
Nsa	1	1	0	0	0	0	1
Nss	1	1	1	0	1	1	1

	X
V	1
Vp	1
Vs	1
Vps	1

	Nid	Nii	Nis	Nia	X
Si	0	0	0	0	1
Sii	1	0	1	1	1
Sid	0	1	1	1	1
Sia	0	0	0	0	1
Sis	0	1	1	1	1

	Noi	Nod	Noa	Nid	Nii	Nis	X
Noi	0	1	1	1	1	1	1
Nod	1	0	1	1	1	1	1
Noa	1	1	0	0	0	0	1
Nos	1	1	1	0	1	1	1

	X
P	0

	Nis	Nia	Nid	Nii	Nis	X
Nii	1	1	1	0	1	1
Nid	1	1	0	1	1	1
Nia	0	0	0	0	0	1
Nis	1	1	0	1	1	1

Tab. 12 : Matrices de compatibilité.

Le tableau 12 présente la table de compatibilité des classes nominales sujet, objet et indirect, de la classe verbale et des classes des particules de type S et P (X désignant n'importe quelle étiquette autre que celles de la ligne 1 ; 0 indique que les étiquettes morpho-syntaxiques peuvent apparaître au sein d'un même tronçon, 1 que non ou que la suite n'est pas attestée en arabe). Par exemple, une frontière de tronçon sépare un sujet déterminé Nsd d'un attribut indéterminé Nsi dans une phrase nominale ; une frontière de tronçon est toujours posée après un verbe (la matrice associée est remplie de 1).

## 5.4. Évaluation

### A. Méthodologie et corpus d'analyse

En raison de l'insuffisance des ressources linguistiques en arabe, nous avons entrepris de constituer nous-mêmes des corpus textuels avec le concours d'un linguiste, bien que cette tâche soit longue et laborieuse. Les structures syntaxiques les plus courantes y sont représentées.

Pour l'évaluation de l'analyseur morpho-syntaxique, nous avons constitué deux corpus de textes : le premier se compose de 123 phrases comptant 1063 mots ; le second de 148 phrases incluant 1169 mots. Les mots de ces deux corpus sont voyellés à l'exception des lettres précédant une pause (c'est le cas généralement pour les textes voyellés). Ils ont été étiquetés manuellement par un expert avec le jeu d'étiquettes que nous avons défini, en indiquant en plus les frontières de tronçons telles que spécifiées.

## B. Résultats

- Corpus 1

L'évaluation du premier corpus a donné un taux d'erreur de 3,1% sur les étiquettes, soit 33 mots mal étiquetés parmi les 1063 mots. Le tableau 13 représente la matrice de confusion des étiquettes.

	V	Vp	Vs	Nsd	Nss	Nsa	Nos	Noa	Nis	Nia	Sia
V	0	1	0	0	0	1	0	1	0	0	0
Vp	0	0	0	0	0	1	0	0	0	0	1
Vps	0	0	1	0	0	0	0	0	0	0	0
Nsi	0	0	0	0	0	1	0	0	0	0	0
Nss	2	0	0	0	0	0	0	0	1	0	0
Nsa	0	0	1	0	0	0	0	0	0	0	0
Noi	1	0	0	0	0	0	0	0	0	0	0
Nos	1	0	1	0	1	0	0	1	1	0	0
Noa	2	0	0	0	0	0	0	0	0	0	0
Nii	0	0	0	0	0	0	0	0	0	1	0
Nid	0	0	0	4	0	0	0	0	0	0	0
Nis	0	0	0	0	0	0	1	0	0	1	0
Nia	2	0	0	0	0	0	0	2	2	0	0
Sis	0	0	0	0	0	0	0	0	1	0	0

Tab. 13 : Matrice de confusion du corpus 1.

## Corpus 2

Pour ce second corpus, nous avons étudié l'apport des règles contextuelles dans le processus d'analyse. Nous avons ainsi effectué deux évaluations : une première évaluation uniquement basée sur l'étiquetage par défaut en cas d'ambiguïté, sans l'application donc des règles contextuelles ; une seconde évaluation avec l'application des règles contextuelles.

### 1- Évaluation sans les règles contextuelles

Une pré-analyse (sans l'application des règles et en gardant les ambiguïtés) a relevé 44 erreurs d'étiquetage et 196 mots ambigus (qui ont plus d'une étiquette). Pour ces derniers, nous avons appliqué trois étiquetages par défaut différents (au cas sujet, au cas objet et au cas indirect) et obtenu les résultats du tableau 14. La lecture de ce tableau indique que l'étiquetage par défaut au *cas indirect* (127 erreurs) donne le meilleur résultat et que l'étiquetage par défaut au *cas direct* (229 erreurs) le moins bon résultat.

Etiquette par défaut	Nombre d'erreurs	Taux d'erreur
Au cas sujet	188	16,08%
Au cas objet	229	19,5%
Au cas indirect	127	10,86%

**Tab. 14 : Résultats de l'étiquetage par défaut aux différents cas.**

## 2- Evaluation avec les règles contextuelles

L'introduction de règles contextuelles a permis un bon étiquetage de 181 mots parmi les mots ambigus ( $x=196-181=15$ ) et la correction de 9 erreurs parmi les 44 mots mal étiquetés ( $y=44-9=35$  erreurs). Le taux d'erreur global est de 4,28%, soit 50 mots mal étiquetés ( $x+y=50$ ). Le tableau 15 représente la matrice de confusion des étiquettes.

	V	Vs	Nsd	Nss	Nsa	Noi	Nos	Noa	Nii	Nid	Nia	Sii	Sia
Nss	1	0	0	0	0	0	0	0	0	0	0	0	0
Nsa	1	0	0	2	0	0	0	0	0	0	0	0	0
Nod	0	0	1	0	0	0	0	0	0	1	0	0	0
Nos	1	2	0	0	0	1	0	0	0	0	0	0	0
Noa	13	0	0	0	3	0	0	0	0	0	0	0	1
Nii	0	0	0	0	0	0	1	2	0	0	0	1	0
Nis	0	0	0	1	2	0	0	0	0	0	4	0	0
Nia	0	0	1	1	13	0	0	2	13	0	0	0	0
Sii	0	0	0	0	0	0	0	0	0	0	0	0	1
Sis	0	0	0	0	0	0	0	0	0	0	0	0	1

**Tab. 15 : Matrice de confusion du corpus 2.**

Il résulte de ces deux évaluations du corpus 2 que, dans le meilleur des cas, l'introduction de règles contextuelles réduit le taux d'erreur de 10,86% à 4,28%. Ce dernier chiffre reste à comparer avec l'état de l'art pour l'anglais et le français, où le taux d'erreur sur les mots tourne autour de 3% [Bou97].

## C. Analyse des résultats

Nous avons recensé les sources d'erreurs les plus importantes pour notre étiqueteur morpho-syntaxique. La répartition de ces erreurs est présentée dans les tableaux 16 (corpus 1) et 17 (corpus 2).

- Les noms à structure verbale

À la lecture de ces tableaux, 12 erreurs dans le premier corpus et 9 dans le second sont dues à des noms qui se présentent sous une forme verbale. Ce type d'erreur peut s'expliquer



par la non-utilisation d'un dictionnaire de racines et l'utilisation d'un traitement se basant essentiellement sur la structure des mots. De plus, l'application des règles contextuelles ne garantit pas toujours la détection des erreurs.

Exemples : Les noms en gras ont une forme verbale

يَحْتَا جُ إِلَى تِلْكَ الْكَانِنَاتِ أَكْثَرَ مِمَّا تَحْتَا جُ إِلَيْهِ  
 نَعْتَمُ إِبْتِي أَنْ نَعُدَّ فُرْصَ حَلْوَى  
 وَتَرَفُضُ الْخُرُوجَ وَحَدَهَا مَا أَنْ يَحْطُ اللَّيْلُ بِظِلَامِهِ

Nom pris pour verbe	Absence de voyelle finale	Verbe pris pour verbe agglutiné ou nom agglutiné	Nom avec /y/ final	Nom indirect pour nom direct	Verbe pris pour nom	Autre
12	2	3	7	2	4	3

Tab. 16 : Répartition des erreurs d'étiquetage du corpus 1.

Nom pris pour verbe	Absence de voyelle finale	Nom propre	Nom avec /y/ final	Nom indirect pour nom direct	Elément de base pris pour affixe
9	15	4	3	5	13

Tab. 17 : Répartition des erreurs d'étiquetage du corpus 2.

- Les mots se terminant par la lettre ي /y/

Ces erreurs sont dues au fait que les signes distinctifs des noms, dans certains cas, sont supposés. Ceci concerne les noms terminés par le *alif* أ /A/ ou *alif maksoura* ى /A/ (الفتى /?alfatA/ « l'enfant »), les noms suivis d'un complément de nom constitué par le pronom lié de la 1<sup>ère</sup> personne du singulier ي /y/ (كتابي /kitAbI/ « mon livre ») et les noms terminés par un ي /y/ précédé d'une *kasra* dans les cas sujet et indirect (القاضي /?alqADI/ « le juge »).

- Les verbes mal étiquetés

Ce type d'erreur concerne principalement les verbes malades : certaines formes verbales, notamment les formes des verbes malades à l'impératif dont la deuxième radicale est doublée, n'ont pas été répertoriées. Il est certain que plus le nombre de formes verbales est important, plus la confusion verbe/nom augmente, ce qui peut affecter les performances du

système. Pour cela, nous avons ignoré les formes à deux lettres (مُدّ /mud/ « donne »), tolérant ainsi volontairement certaines erreurs sur les verbes. Ce type d'erreur, comme nous le verrons plus loin, a des incidences négligeables sur les frontières des tronçons.

- Les mots dont les éléments de base sont assimilés à des préfixes ou suffixes

Dans un mot, lorsqu'un élément de base initial ou final s'apparente à un préfixe ou à un suffixe, celui-ci est considéré comme un mot composé (préfixe + base ou base + suffixe). Par exemple, l'élément de base كَ /ka/ de كَبِير /kabIrin/ (« grand ») est pris pour un préfixe ; l'élément de base ان /Ani/ de اِثْنَان /?icnAni/ (« deux ») est pris pour un suffixe.

- Erreurs dues aux noms propres

Certains noms propres sont invariables quelle que soit leur position dans la phrase. Leur signe distinctif ne suit pas leur fonction, ce qui constitue une source d'erreur pour un analyseur se basant sur la flexion casuelle des mots. Exemple : نَجَّحَ مِنْ بَيْنِهِمْ عُمَرَ وَ مَهْدِي. Le sujet مهدي /mahdI/ du verbe نجح /najaHa/ est étiqueté Nia (nom indirect).

- L'absence de la voyelle finale

À l'instar des noms propres, les mots dont la dernière lettre n'est pas voyellée sont mal étiquetés. Exemple : مُتَعَدِّدَةٌ مَيَّادِينِ (Si) فِي أَسْهَمَ. Le mot مُتَعَدِّدَةٌ /mutæaddida/ (« multiple ») est étiqueté Nia (nom au cas indirect) au lieu de Nii (nom indirect indéterminé) en raison de l'absence de sa voyelle finale. Il faut noter que l'analyseur a été initialement conçu pour traiter des textes entièrement voyellés. Néanmoins, son application sur les corpus 1 et 2 nous a renseigné sur sa capacité à s'adapter à des situations réelles, où les voyelles finales des phrases sont omises.

#### D. Incidence sur les frontières de tronçons

Les erreurs d'étiquetage ont un impact variable sur les frontières de tronçons. Ainsi, seules 11 erreurs sur 33 ont une incidence sur le découpage en tronçons (33%) en ce qui concerne le premier corpus, et 7 sur 50 en ce qui concerne le second (14%).

- Exemple d'erreurs avec incidence sur les tronçons

Dans ces exemples, les crochets représentent les frontières théoriques et les parenthèses les frontières obtenues après analyse automatique.

[ مَرَّةً ( أُخْرَى ) ] فَإِنَّ الْإِنْسَانَ يُحَطِّمُ مَلَادَّ الْكَائِنَاتِ  
 لَمْ يَمْضِ عَلَى [ تَأْسِيسِ ( بَعْدَادِ ) ] وَقَتَّ طَوِيلٌ  
 مِمَّا يَزِيدُهَا جَمَالًا أَنْ فِيهَا حَدَائِقُ ( كَثِيرَةٌ )

- Exemple d'erreurs sans incidence sur les tronçons

يَحْتَاجُ إِلَى تِلْكَ الْكَائِنَاتِ [ ( أَكْثَرِ ) ] مِمَّا تَحْتَاجُ إِلَيْهِ  
 وَ تَرْفُضُ الْخُرُوجَ [ ( وَحْدَهَا ) ] مَا أَنْ يَحُطُّ اللَّيْلُ بِظِلَامِهِ  
 تَمَامًا [ ( فَتْرَاهُ ) ] بِصَوْتِهِ الْغَلِيظِ وَ حَرَكَاتِهِ الضَّخْمَةِ

## 5.5. Discussion

La finalité de l'analyseur morpho-syntaxique est l'étiquetage des mots en vue de leur regroupement en tronçons. À partir de là, les performances de cet analyseur ne doivent pas être observées au niveau du mot, mais au niveau du tronçon. Ce postulat a orienté la méthodologie de segmentation adoptée dans le sens où, volontairement, certaines erreurs d'étiquetage sont tolérées si elles n'influent pas sur la segmentation en tronçons. Une erreur d'étiquetage n'est donc *critique* que si elle est susceptible d'introduire ou de supprimer une frontière de tronçon.

	Nom pris pour verbe	Absence de voyelle finale	Nom propre	Nom avec /y/ final	Elément de base pris pour affixe
Corpus1	5	1	0	5	0
Corpus2	1	0	3	2	1

Tab. 18 : Répartition des erreurs.

Le tableau 18 représente la répartition des incidences sur les frontières de tronçons par classe d'erreurs. Ainsi, bien que la liste des formes verbales soit incomplète (il manque les verbes malades à l'impératif), l'incidence d'une confusion verbe/nom sur une frontière de tronçon reste peu probable (aucune selon les résultats du tableau). Les erreurs les plus fréquentes sont celles occasionnées par la confusion nom/verbe, les noms propres et les mots se terminant par la lettre /y/, les autres erreurs ayant une incidence minimale.

Pour pallier ces erreurs, le recours à un dictionnaire de mots lexicaux et de noms propres (les plus courants dans la langue) s'avère nécessaire, sans exclure la possibilité d'affiner et d'enrichir la base de règles contextuelles utilisée. Pour le premier cas, et au vu de la difficulté de la mise en œuvre de nouvelles ressources linguistiques, nous pensons que le gain de performance recherché ne justifie pas un tel investissement.

## CHAPITRE 6 : TRANSCRIPTION ORTHOGRAPHIQUE-PHONÉTIQUE, SYLLABATION ET ACCENTUATION

Nous présentons dans ce chapitre notre étude sur la transcription orthographique-phonétique de textes arabes voyellés en vue de la SAT. Nous décrivons ainsi les approches utilisées dans ce domaine et survolerons les difficultés que peut poser l'automatisation de cette tâche en arabe. Nous détaillerons enfin le système que nous avons développé.

### 6.1. *Transcription orthographique-phonétique*

La transcription orthographique phonétique, ou phonétisation, est une étape essentielle dans un système de SAT. Elle consiste à produire la prononciation correspondant au texte en entrée sous la forme d'une liste de phonèmes. Dans certains cas, cette séquence phonétique inclut des marqueurs symboliques liés à l'accentuation et à l'intonation du texte (cf. section 6.2).

Généralement, la TOP intervient après les pré-traitements et l'analyse syntaxique. Sa complexité varie selon la langue traitée et la nature du texte en entrée. Par exemple, la langue française est très difficile à transcrire en raison de sa forme orthographique qui est très différente de sa forme phonétique (les noms propres posent de plus un problème complexe car leur prononciation dépend fortement de leur origine), contrairement à l'arabe où la correspondance entre les graphèmes et les phonèmes est quasi-biunivoque.

Dans la pratique, les textes comportent beaucoup d'anomalies, c'est-à-dire, des mots qui ne sont pas (ou peu) représentés dans le dictionnaire. Parmi eux, les nombres, les sigles, les symboles et autres abréviations. La TOP doit prendre en compte ces problèmes en apportant si nécessaire des solutions spécifiques : il convient de savoir si les sigles doivent être épelés ou non [Bou97].

#### 6.1.1. Approches en transcription orthographique-phonétique

Nous exposons brièvement dans ce qui suit les approches utilisées pour la TOP dans les différentes langues. Yvon [Yvo96] et Boula de Mareuil [Bou97] présentent une lecture plus approfondie de ce sujet. La TOP peut se faire grâce à l'utilisation de lexiques et/ou de règles de réécriture.

##### A. Approches à base de lexiques

Dans cette approche, la phonétisation se fait grâce à un lexique qui associe à une entrée lexicale, sa forme phonétisée. Souvent, les dictionnaires électroniques ne comportent que les formes *canoniques* des mots, ce qui nécessite le recours à un traitement (des règles) permettant de déduire la forme phonétique des mots à partir de leur forme fléchie [Sar90].

L'avantage des approches à base de lexiques est la souplesse dans la maintenance et la mise à jour des connaissances. De plus, la disponibilité des supports de stockage aux capacités toujours croissantes et l'implémentation d'algorithmes de recherche et d'accès aux lexiques de plus en plus rapides favorisent *incontestablement* le recours à de telles approches.

## B. Approche à base de règles

Dans cette approche, les connaissances sont spécifiées par des *règles de réécriture* qui convertissent des graphèmes en phonèmes selon leurs contextes gauche et droit. Néanmoins, pour traiter les nombreux cas particuliers, un dictionnaire d'exceptions est nécessaire : en anglais, il faudra environ 500 règles de transcription et un lexique de 7000 exceptions pour réaliser une transcription correcte des 20.000 mots les plus courants. Par contre, en espagnol, 50 règles sont suffisantes pour effectuer une bonne transcription [Mor98].

L'avantage de cette approche est l'utilisation d'un formalisme de description des connaissances proche de celui employé en linguistique. Parmi les langages de programmation par règles de réécriture, citons `flex` et `Compost` [Ali93]. Ce dernier est utilisé pour la TOP dans le système de SAT du français développé à l'ICP<sup>12</sup>. Sa syntaxe est la suivante :  $C \rightarrow R / G + D$ . Elle se lit : C (la cible) se réécrit R (remplacement) si elle est précédée de G (contexte gauche) et suivie de D (contexte droit).

L'inconvénient de l'approche à base de règles réside dans la gestion de la quantité de règles implémentées. En effet, dès lors que leur nombre devient trop élevé, il devient difficile d'introduire de nouvelles règles sans occasionner de situation conflictuelle : si deux règles (ou plus) s'appliquent pour un graphème (ou groupe de graphèmes) dans un contexte donné, alors la règle la plus proche du début est appliquée dans `Compost` ; la règle qui renferme le plus grand nombre de graphèmes, sinon la première règle rencontrée dans `flex`, ce qui nécessite une organisation minutieuse de la base. Morel [Mor98] quant à lui, propose une structuration arborescente des règles pour la TOP du français.

La plupart des systèmes de TOP actuels combinent les règles et les lexiques en proportion variable. Selon les difficultés de la langue traitée, la priorité peut être donnée aux lexiques (l'anglais se prête mieux à une approche par lexiques [Boi00]) ou aux règles (l'espagnol [Mor98]). L'arabe appartient à cette deuxième catégorie de langues en raison de la correspondance biunivoque qui existe entre les graphèmes et les phonèmes, à quelques exceptions près [Bal02a] [Sar90].

### 6.1.2. Systèmes de transcription de textes arabes

Nous allons passer en revue quelques systèmes de TOP de l'arabe standard. L'ensemble de ces systèmes se distinguent par les connaissances linguistiques mises en œuvre pour la

---

<sup>12</sup> Institut de la Communication Parlée - Grenoble

génération de la chaîne phonétique et par la manière dont sont organisées et utilisées ces connaissances. Ils se caractérisent par l'utilisation à la fois de lexiques (d'exceptions ou généraux) et de règles de conversion.

- Les travaux de ZEMIRLI

Le système SYNTHAR+ [Zem98a] développé à l'Institut National d'Informatique d'Alger (INI) génère d'abord la chaîne phonétique d'un texte arabe voyellé à partir de sa représentation graphique, puis la transmet au synthétiseur Multivox de la société hongroise AKADIMPEX pour la génération acoustique. Une analyse morphologique vérifie d'abord l'appartenance des mots à la langue et, le cas échéant, leur attribue des valeurs morphologiques et syntaxiques. Ces informations sont utilisées pour la TOP qui repose sur l'utilisation de règles contextuelles pour les traitements morpho-orthographiques et phonologiques, et sur l'utilisation de lexiques d'exceptions et d'un lexique général pour les traitements morpho-lexicaux.

- Les travaux de SAROH

Dans le cadre du projet SYAMSA (SYstème d'Analyse Morpho-Syntaxique de l'Arabe), Saroh [Sar90] a développé un convertisseur graphème-phonème pour l'arabe voyellé. Selon lui, « la phonétisation de l'arabe repose en particulier sur l'emploi de lexiques et d'un analyseur morphologique pour la génération des différentes formes d'un mot. Par ailleurs, ce sont les phénomènes d'interaction entre les mots (liaison, élision, etc.) et les phénomènes d'assimilation qui suggèrent l'utilisation de règles phonologiques ».

Ce système repose sur l'utilisation d'une base lexicale relationnelle. Ainsi, chaque entrée lexicale inclut la représentation graphique d'une racine, ses attributs morphologiques (type de la racine, catégorie grammaticale, etc.) ainsi que la forme phonétique correspondante. Des règles sont implémentées pour déduire les formes dérivées et fléchies des racines et les descriptions phonétiques qui leur sont associées. Dans un premier temps, la TOP consiste à comparer chaque mot dans le texte à l'ensemble des mots issus de la génération morphologique et, en cas de succès, à fournir la séquence phonétique correspondante.

Comparé au système de ZEMIRLI, ce système intègre une composante syntaxique qui, par une méthode ascendante, valide la structure de la phrase à partir de traits morpho-syntaxiques associés aux mots. Aussi, l'analyse syntaxique permet de positionner des marqueurs de pause ou de liaison dans la phrase, ce qui conditionne le déclenchement de règles phonologiques pour le traitement des phénomènes d'interaction entre les mots (liaison, assimilation, etc.).

- Les travaux de GHAZALI

Ghazali [Gha92b] a développé un système de TOP à l'Institut Régional des Sciences Informatiques et des Télécommunications (IRSIT) de Tunis en vue de la SAT arabe voyellée. Ce système se distingue des deux premiers par le niveau de connaissances phonologiques implémentées. Ainsi, après les pré-traitements du texte, la conversion graphème-phonème et l'application de règles phonologiques d'interaction entre les mots, des règles de *propagation de l'emphase* sont employées.

### 6.1.3. Difficultés en transcription graphème-phonème de l'arabe

Nous classerons dans ce qui suit les difficultés de TOP de textes arabes en deux catégories, selon qu'elles apparaissent à l'intérieur du mot ou à ses frontières. Nous ne parlerons pas des problèmes liés aux mots extra-lexicaux (sigles, dates, symboles, etc.) dont la résolution relève de traitements indépendants de la langue ni de la problématique de phonétisation de textes non voyellés, car l'absence de voyelles entraînerait inévitablement une ambiguïté de la prononciation des mots.

#### A. Difficultés intra-mot

- La première difficulté des systèmes de TOP dans les différentes langues découle du fait que la correspondance « un graphème-un phonème » n'est pas toujours respectée. L'arabe n'échappe pas à cette règle :
  - Un graphème peut correspondre à des *phonèmes* différents : les trois graphèmes و ي correspondent soit à des voyelles longues, respectivement /A/ /U/ /I/, s'ils ne portent pas de voyelles brèves, soit à des consonnes, respectivement /ʔ/ /w/ /y/ s'ils portent une voyelle brève. Par exemple, le graphème ي est transcrit /I/ dans le mot رِيح /rIHun/ (« vent ») et /y/ dans يَدُهُ /yadah/ (« sa main »).
  - Un graphème peut correspondre à plusieurs phonèmes : c'est le cas des trois graphèmes du tanwin (ّ) qui correspondent respectivement aux phonèmes /an /un/ /in/ et du graphème de la *madda* َ qui correspond aux phonèmes /ʔA/. Exemples : مَنْزِلٌ /manzilun/ (« maison »), آلَةٌ /ʔAlatun/ (« machine »).
  - À l'inverse, plusieurs graphèmes peuvent correspondre à un seul phonème. C'est le cas des graphèmes ت et ة qui correspondent au phonème /t/, des graphèmes ؤ و ئ أ ة qui correspondent au phonème /ʔ/, et des graphèmes ا ي qui correspondent au phonème /A/ (si les deux graphèmes dans ce dernier cas ne suivent pas les *fathatan* en fin de mot et si le premier d'entre eux n'est pas signé). Exemples : جَائِزَةٌ /jAʔizata/ (« prix »), تَقْرَأُ /taqraʔu/ (« elle lit »), تُؤْمِرُ /tuʔmiru/ (« elle ordonne »), عَلَى /ealA/ (« sur »), قَالَ /qAla/ (« il a dit »). À l'exception des cas cités ci-dessus, l'ensemble des autres graphèmes ont une correspondance phonémique unique.

- Le traitement des graphèmes correspondant au *son zéro*. Ce sont des graphèmes qui, bien que présents à l'écrit, ne se prononcent pas dans certains contextes. C'est le cas des graphèmes /ʔ/ et /ʕ/ après les *fathatan* en fin de mot et du graphème /ʔ/ en fin de verbe au pluriel masculin. Exemples : غَدًا /gadan/ (« demain »), فَتَى /fatan/ (« enfant »), دَخَلُوا /daxalU/ (« ils sont entrés »). Le graphème ʔ correspondant au *soukoun* n'est quant à lui jamais prononcé.
- Le traitement de la *chadda*. La présence de la *chadda* se traduit par le dédoublement de la consonne qui la porte. Exemple : تَسَلَّمَ /tasallama/ (« il a reçu »).
- Le traitement des mots irréguliers. Ce sont des mots dont la prononciation ne correspond pas exactement à leur graphie, sans qu'elle soit régie par des règles de conversion attestées :
  1. Le graphème /A/ est absent à l'écrit mais prononcé à l'oral. Exemples : اللَّهُ /ʔallAhu/ (« dieu »), ذَلِكَ /vAlika/ (« celui-là »), لَكِنَّ /lAkinna/ (« sauf que »).
  2. Le graphème /A/ est présent à l'écrit mais ne se prononce pas. Exemples : مائة /miʔatun/ (« cent »), مائتان /miʔatAni/ (« deux cents »).
  3. Le graphème /ʔ/ est déplacé à l'oral. Exemples : هَذَا /hAva/ (« ce »), هَكَذَا /hAkava/ (« comme ça ») (le graphème /ʔ/ est déplacé vers le début du mot).
  4. Le graphème w /U/ est présent à l'écrit mais ne se prononce pas. Exemples : أُولَئِكَ /ʔulAʔika/ (« ceux-là »), عَمْرُو /eamrun/, بَعْمَرُو /biamrin/ (ces deux derniers exemples correspondent à des noms propres). À l'inverse, le graphème و /U/ se prononce dans certains noms propres bien qu'absent à l'écrit. Exemples : دَاوُدَ /dAwUda/.
- Le traitement de la propagation de l'emphase. Rappelons que l'emphase concerne les voyelles brèves et longues qui changent de timbre au voisinage des consonnes emphatiques ص /S/ ط /T/ ظ /Z/ ض /D/. La complexité de ce problème vient du fait que l'emphase se propage aux phonèmes voisins, rendant difficile de déterminer le segment affecté par cette propagation.
- Le traitement des nombres. La phonétisation d'un nombre dépend de sa *fonction* (cas sujet, objet ou indirect) dans la phrase d'une part et du *genre* (masculin/féminin) du nom auquel il est rattaché d'autre part. Par exemple, (« 5 رجال ») est prononcé (« خمسة رجال ») /xamsatu rijAlun/ (« cinq hommes ») (رجال est masculin) ; (« 5 نساء ») est prononcé (« خمس نساء ») /xamsu nisAʔ/ (« cinq femmes ») (نساء est féminin), (« في 5 مراحل ») est prononcé (« في خمسة مراحل ») /fi xamsati marAhila/ (« en cinq étapes »). Dans ce dernier cas, le chiffre 5 est précédé de la préposition في /fi/ qui le met au cas indirect (flexion du cas indirect). Une analyse morphologique s'avère donc nécessaire pour la TOP des nombres en arabe.



## B. Difficultés extra-lexicales

La transcription d'unités de taille supérieure au mot engendre de nouveaux problèmes liés aux phénomènes d'interaction entre les mots. Ces interactions conduisent souvent à la suppression/ajout de graphèmes aux frontières des mots, sur la base de connaissances phonologiques et syntaxiques de la langue.

- La première difficulté en phonétisation d'une séquence extra-lexicale est liée au phénomène d'*assimilation* qui peut intervenir de différentes manières :
  1. Il intervient d'abord entre deux consonnes identiques situées aux frontières des mots, la première portant la *soukoun* ° et la seconde une voyelle brève. L'assimilation se traduit alors par le dédoublement de la consonne en question. Exemple : قُلْ لَهُ /qul lahu/ (« dis-lui ») se prononce قُلُّهُ /qullahu/.
  2. Il intervient au niveau de l'article « ال » qui, selon son contexte, se prononce de différentes manières. En début de phrase : si l'article est suivi d'un phonème *lunaire*, il est alors prononcé /ʔal/ (Ex : الولد /ʔalwaladu/ « l'enfant ») ; si l'article est suivi d'un phonème solaire, alors le phonème ل /l/ est supprimé (Ex : السَّمَاء /ʔassamAʔu/ « le ciel »). En milieu de phrase : si l'article /ʔal/ est suivi d'un phonème *lunaire*, le phonème /ʔ/ est alors supprimé (Ex : في الأرض /fil ʔarDi/ « sur la terre ») ; si l'article est suivi d'un phonème solaire, il est alors supprimé (Ex : في السَّمَاء /fissamAʔi/ « dans le ciel »).
- Certains phonèmes sont modifiés aux frontières des mots. C'est le cas des voyelles longues /A/ /U/ /I/ qui se transforment en voyelles brèves, respectivement /a/ /u/ /i/, si elles sont suivies de l'article ال. Exemples : عَلَى الطَّوَلَةِ /ealaT TAWilati/ (« sur la table »), أَكَلَا الخُبْزَ /ʔakalal xubza/ (« ils ont mangé du pain (au duel) »).
- Certains phonèmes sont ajoutés à la frontière des mots dans certains contextes. Ainsi, si une consonne portant la *soukoun* en fin de mot est suivie de l'article ال, elle prend alors la *damma*. Exemple : عَلَيْكُمْ السَّلَامَ /ealaykumus salAm/ (« que la paix soit sur vous ») (le phonème /u/ est introduit à la fin du premier mot /ealaykum/). L'exception à cette règle est le graphème ن /n/ de la préposition مِنْ /min/ qui prend une *fatha*. Exemple : مِنَ الأسبوعِ /minal ʔusbUei/ (« de la semaine ») (le phonème /a/ est introduit à la fin du premier mot /min/).
- Le traitement de la *hamza* constitue une autre difficulté en phonétisation de séquences extra-lexicales. Celle-ci est soumise à diverses déformations déterminées par son contexte d'apparition. Ainsi, elle est prononcée en début de phrase (Ex : أَكْرِمُ /ʔakrim/ « fais du bien »), mais peut être omise à son milieu (Ex : رَأَيْتُ ابْنَهَا /raʔaytub nahA/ « j'ai vu son fils ») où la hamza en début du deuxième mot est supprimée). Divers travaux lui ont consacré un traitement spécifique de part son importance dans la langue [Sar90].

#### 6.1.4. Développement d'un phonétiseur pour l'arabe standard voyellé

Dans cette section, nous présentons le module de TOP que nous avons développé pour l'arabe standard voyellé et intégré dans le système d'Elan Speech [Bal02a]. Ce système, rappelons-le, synthétise des documents de nature différente (email, fax, etc.), d'où la nécessité d'une phase de *normalisation* du texte en entrée avant la TOP proprement dite.

##### A. Normalisation du texte

La *normalisation* du texte a pour mission de transformer un texte pouvant contenir des anomalies en un texte *normalisé* dont l'ensemble des mots est transcrit en toutes lettres. Elle se déroule en deux phases. Dans la première phase, le texte est découpé en phrases structurées en mots. Dans la deuxième phase, certaines expressions (unités de mesure, monétaires, abréviations, nombres, etc.) sont réécrites en toutes lettres.

##### 1. Tokenisation

Le texte en entrée est filtré par un repérage des frontières de mots et de phrases (le mot ici est pris au sens large) sur la base des indicateurs de surface comme le blanc, le saut de ligne ou plus généralement des marques de ponctuation. Cette opération n'est pas toujours simple en raison des interférences que ces marques peuvent engendrer. Par exemple, le point peut conclure une phrase, apparaître dans un sigle, séparer les dizaines des centaines dans un nombre (100.000), etc. Par ailleurs, les mots alphanumériques sont séparés en chiffres et en lettres (95 وندوز « windows95 » est séparé en وندوز et 95). Il en résulte à l'issue de la *tokenisation* une séquence de mots associée à chaque phrase.

##### 2. Pré-traitements

Les pré-traitements ont pour tâche de transcrire en toutes lettres l'ensemble des mots, en particulier les mots extra-lexicaux comme les nombres et les abréviations. Ces problèmes doivent être résolus pour la synthèse vocale à partir de textes [Bou97] [Yvo98]. Le tableau 19 résume l'essentiel de ce que les pré-traitements prennent en compte pour les langues implémentées dans le système d'Elan Speech.

Dans la colonne du milieu figure le type de motif (numéroté de 1 à F en hexadécimal). Les colonnes qui l'entourent donnent des exemples (comme le point) de caractères spéciaux entrant dans la composition de chacun.

#	espace ou tabulation	type de motif	caractère
1	blanc	abréviations communes	tiret arobase point virgule deux-points slash apostrophe
2		prénoms et titres honorifiques	
3		sigles avec points abrégatifs	
4		dates, heures et durées	
5		numéros de téléphone	
6		monnaies	
7		nombres décimaux, ordinaux...	
8		chiffres romains	
9		URL et adresses électroniques	
A		mots pouvant être épelés	
B		fractions	
C		unités de mesure (SI...)	
D		expressions mathématiques	
E		symboles non alphanumériques	
F		extensions de fichiers info.	

Tab. 19 : Pré-traitements des textes dans le système de SAT d'Elan Speech.

La capture des motifs est réalisée à l'aide de `flex`, un générateur d'analyseurs lexicaux. Il permet d'exprimer les chaînes à manipuler ou *motifs* sous forme d'*expressions régulières* de manière déclarative. Ces spécifications sont ensuite transformées en programme C (ou C++) optimisé qui s'intègre de manière efficace dans l'architecture SYC du système d'Elan Speech. Le langage défini par `flex` recouvre exactement celui qu'on peut décrire avec les *automates* (non déterministes, à partir desquels on peut choisir des automates déterministes). Des exemples de motifs pour les dates et les sigles sont :

$$([01]?[0-9]|2[0-4]):[0-5][0-9]$$

$$(\{\text{consonne}\}\.){2,5}$$

où `{consonne}` désigne la classe de consonnes (classe définie par le programmeur) et où les parenthèses délimitent les arguments du « ou » (`()`). Pour ce dernier exemple, nous tolérons des séquences de 2 à 5 consonnes pour les sigles.

Un programme `flex` cherche pour la fenêtre de texte en entrée le motif le plus approprié parmi l'ensemble des motifs définis. Si plusieurs motifs sont applicables, alors le motif qui consomme le plus d'éléments est sélectionné. Puis, si l'ambiguïté est maintenue, c'est le premier motif dans l'ordre des spécifications des motifs qui est retenu.

- Traitement des sigles et abréviations

Un sigle en arabe est une forme abrégée construite en concaténant les consonnes initiales d'une séquence de mots (en excluant les clitiques). Ex : التلفزيون = ت.و.ج. الوطني الجزائري (« la télévision nationale algérienne »). Une abréviation consiste à réduire un mot ou groupe de mots en n'utilisant qu'une partie de ses lettres (Ex : الخ =

إلى آخره (« etc »)). Les sigles et abréviations sont stockés dans des lexiques généraux ou propres au domaine d'utilisation. Par exemple, le sigle !.و.م.ا. يمكن أن يكون له معنيين محتملين :

- a. المَدْرَسَةُ الوَطَنِيَّةُ لِلإِدَارَةِ : م.و.ا. (« école nationale d'administration ») dans le domaine de l'enseignement supérieur ;
- b. المُؤَسَّسَةُ الوَطَنِيَّةُ لِلإِنْتِاجِ : م.و.ا. (« l'entreprise nationale de production ») dans le domaine économique.

Le traitement consiste à comparer les mots dans le texte à l'ensemble des mots dans les lexiques et à les remplacer par leur traduction en cas de succès.

#### ▪ Traitement des dates et des heures

Les expressions qui respectent les syntaxes aaaa/mm/jj ou aaaa.mm.jj déclenchent les procédures de traitement des dates. Exemple : 2000/012/01 est traduit par أول ديسمبر ألفين ; les expressions ayant la syntaxe hh:mm:ss déclenchent les procédures de traitement des heures. Exemple : 1 : 15 est traduit par الواحدة و أربع دقائق.

#### ▪ Traitement des nombres

Comme nous l'avons mentionné, la phonétisation d'un nombre se fait selon sa fonction dans la phrase et selon le genre (masculin/féminin) du mot auquel il est rattaché. Nous avons de ce fait implémenté des heuristiques sous forme de règles pour la détermination de ce trait morphologique (exemple d'une heuristique : un nom se terminant par un /t/ ou /At/ est de genre féminin). Ainsi, 13 سَيَالَةٍ (« 13 stylos ») est traduit par /calAcu eaorapa sayyAlatin/ (nom féminin), 3 كُتُبٍ (« 3 livres ») est traduit par /calAcatu kutubin/.

#### ▪ Traitement des unités spéciales

Le traitement ici concerne les *motifs* renfermant les unités de mesures, de monnaies, etc. Exemple : 5م.م. est traduit par /xamsapu ?amtAr/. خمسة أمتار.

### 3. Transcription orthographique-phonétique

Le texte « en toutes lettres » est présenté au module de TOP à l'issue des pré-traitements et de l'analyse syntaxique (cf. chapitre 5). La TOP de l'arabe repose sur une centaine de règles et un lexique d'exceptions (cf. figure 14). Elle se déroule comme suit :

- Traitement des mots irréguliers

Les mots irréguliers sont répertoriés dans un lexique d'exceptions de 20 entrées qui fournit en sortie les formes phonétiques. Rappelons que la prononciation de ces mots ne correspond pas à leur graphie. Exemples : le mot هَذَا /havA/ est transcrit /hAva/ ; le mot ذَلِكَ /valika/ est transcrit /vAlika/.

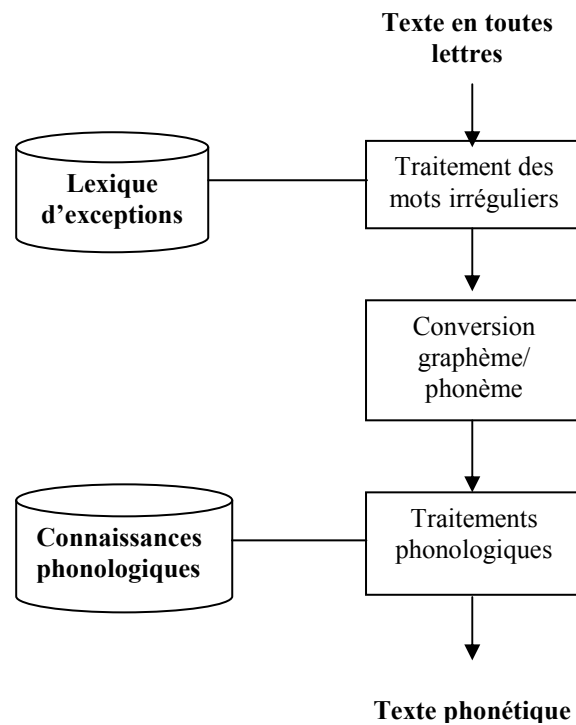


Fig. 14 : Module de transcription orthographique-phonétique.

- Conversion graphème/phonème

Cette deuxième étape opère la conversion proprement dite. Chaque graphème ou groupe de graphèmes est remplacé par un ou plusieurs phonèmes selon ses contextes gauche et droit. Les règles de conversion rendent compte des exceptions *morpho-orthographiques*, du *tanwin*, de la *chadda*, de la *madda* et des sons *Zéro*. Des exemples de ces règles sont (en adoptant la syntaxe flex):

- [ و ] --> /U/. Ex : دَخَلُوا --> /daxalU/ (« ils sont entrés ») (le و /w/ du verbe masculin au pluriel ; où [ ] désigne un espace) ;
- [ ي ] --> /an/. Ex : -فتى- --> /fatan/ (« enfant ») ;

- ء--> /an/. Ex : مَدْرَسَةٌ --> /madrasapun/ (« école ») (le *tanwin*) ;
- آ--> /ʔA/. Ex : أَفَاقٌ --> /ʔAfAq/ (« annales ») (le *madd*) ;
- {c} --> {c}{c}. Exemple : كَسَّرَ --> /kassara/ (« il a cassé ») (la *chadda* ; où {c} désigne n'importe quelle consonne) ;
- [تت] --> /t/. Exemple : تَارَةً --> /tAratun/ (« quelques fois ») (où [xy] désigne x ou y).

- Traitements phonologiques

Les règles phonologiques rendent compte des phénomènes de transformation de la *hamza*, de la liaison, de l'assimilation et de l'emphase.

- Traitement de la *hamza*

La *hamza* a fait l'objet d'un traitement spécifique de part son importance dans la langue. Nous avons ainsi élaboré 15 règles qui rendent compte de son contexte d'apparition. Certaines de ces règles sont obligatoires, d'autres facultatives. Des exemples de ces règles sont :

- [أ]^ --> /ʔ/. Exemple : ابْنُهُ --> /ʔibnuhu/ (« son fils » - la *hamza* se prononce en début de phrase ; où ^ désigne le début de la chaîne) ;
- [أ][ ] --> /a/. Exemple : مَعَ ابْنِهِ : --> /macabnihi/ (« avec son fils » - la *hamza* est précédée d'une *fatha*, elle est retranchée).

- Règles d'assimilation

Les règles d'assimilation rendent compte des transformations de l'article ال/ʔal/ selon sa position dans la phrase et la nature du phonème (lunaire/solaire) qui le suit. Des exemples de ces règles sont :

- {lunaire}ال^ → /ʔal/. Exemple : البُسْتَانُ → /ʔalbustAnu/ (« le jardin ») (l'article se trouve en début de phrase suivi d'une consonne lunaire) ;
- {solaire}ال^ → /ʔa/{solaire}{solaire}. Exemple : السَّمَاءُ → /ʔassamAʔu/ (« le ciel ») (l'article se trouve en début de phrase suivi d'une consonne solaire) ;
- {lunaire}ال[ ] → /al/. Exemple : فَوْقَ [ ] الْمَكْتَبِ → /fawqal maktabi/ (« sur le bureau ») (l'article se trouve en milieu de phrase, précédé d'une *fatha* et suivi d'une consonne lunaire) ;
- {solaire}ال[ ] ي → /i/{solaire}{solaire}. Exemple : فِي [ ] السَّمَاءِ → /fissamAʔi/ (« dans le ciel ») (l'article se trouve en milieu de phrase, précédé d'une voyelle longue et suivi d'une consonne solaire) ;

## – Traitement de l'emphase

Les règles de l'emphase implémentées rendent compte des voyelles au contact des consonnes emphatiques ص ض ط ظ. La propagation de l'emphase quant à elle n'est pas modélisée. Des exemples de ces règles sont :

- ض → /Tâ/. Exemple : ضَرَبَ → /Dâra/ (« il a frappé »). (la voyelle se trouve après la consonne ; où /â/ désigne la variante emphatique de /a/).
- [ظ ط ص ض] → /â/. Exemple : البَصْرُ → /?albâSâr/ (« la vue ») (la voyelle se trouve avant la consonne).

### 6.1.5. Discussion

Cette discussion porte sur divers points comparés avec le français :

#### ○ La transcription des éléments extra-lexicaux

Nous comptons parmi eux les noms propres et les mots d'emprunt qui posent, selon les langues, des difficultés très variables. La phonétisation des noms propres en arabe, quand ils sont isolés, ne pose pas de problème particulier du fait que leur forme orthographique est proche de leur forme phonétique. Par contre, la situation est différente à l'intérieur des phrases ou des noms composés : si un nom propre compte un mot déterminé par l'article ال, l'assimilation de cet article n'est alors pas systématique. Par exemple, dans le nom composé رافع الطهطاوي /RkFIε uaTTahTAwI/, l'article ال /?al/ de الطهطاوي /uaTTahAwI/ est préservé. La résolution de ce problème passe par la définition d'un lexique des noms propres ambigus.

#### ○ La transcription des mots d'emprunt

Ces mots ont une origine différente des mots arabes (langue indo-européenne). Une première phase de translittération est donc nécessaire pour passer de l'alphabet d'origine vers l'alphabet arabe. Les mots produits sont facilement prononçables dans la mesure où ils sont proches de la phonétique. Cependant, l'origine de ces mots nécessite souvent un effort au niveau de l'adaptation : certaines langues, comme les langues européennes, ont un système phonétique différent de celui de l'arabe. Ce qui peut donner dans certains cas une prononciation sans rapport avec les sons d'origine. Par exemple, la séquence « Société Générale », après une translittération en « سوسيتي جنرال » (journal El-khabar du 23/01/02) est prononcée /susyiti jiniral/.

#### ○ La transcription des sigles

Il existe différentes manières de lire un sigle en français, les deux principales étant la lecture et l'épellation [Bou97]. Rappelons qu'un sigle en arabe se compose de consonnes séparées par des points. L'absence de voyelles rend leur lecture impossible et offre comme

seule solution leur *épellation*, à supposer qu'ils soient absents des lexiques qui leur sont dédiés.

- La transcription des *homographes hétérophones*

Ces mots se prononcent de manières différentes selon leur catégorie ou leur sens dans la phrase. Ils sont fréquents en langue française, ce qui peut occasionner des problèmes dans le processus de phonétisation. Par exemple, les mots « est » et « couvent » se prononcent différemment selon qu'ils sont de nature verbale ou nominale (le ciel est bleu, le soleil se lève à l'Est, les poules du couvent couvent).

A priori, ce problème n'existe pas en arabe standard voyellé. Les mots se prononcent de façon unique quel que soit leur contexte dans la phrase. Néanmoins, il peut y avoir des ambiguïtés de prononciation dans les dialectes régionaux. Harkat [Har92] cite un cas d'ambiguïté dans le *parlé algérien* qui concerne le mot راب. Celui-ci est transcrit selon son sens dans la phrase : راب الحليب /rAbal HalIbu/ (« le lait a caillé ») (voyelle non-emphatisée sur la première syllabe) et راب الحيط /rÂbal HITû/ (« le mur s'est écroulé ») (voyelle emphatisée sur la première syllabe). Cet exemple constitue à notre connaissance le seul exemple dans le *parlé algérien* et n'est pas observé en arabe *standard*.

## 6.2. SYLLABATION ET ACCENTUATION

### 6.2.1. Généralités

Le terme *accentuation* représente le phénomène de *mise en relief* de certaines syllabes par rapport aux syllabes voisines, ce qui leur procure une perception plus forte que leur entourage. Le terme *proéminence* (actualisation et perception) est également utilisé dans la littérature. Le domaine de l'accentuation est le mot, d'où les appellations *accent de mot* et *accent lexical*.

Certaines langues sont dites à *accent fixe* en raison de l'accent lexical qui se manifeste **toujours** à une position fixe dans le mot (la première syllabe pour la langue tchèque, la dernière pour la langue française). Les langues dans lesquels la position de l'accent lexical est variable (comme l'anglais et l'italien) sont dites à *accent libre*. Dans ce cas, sa fonction peut être *distinctive* : par exemple, le mot « ancora » en italien signifie « ancre » si l'accent se trouve sur la première syllabe (**an**cora), et « encore » s'il se trouve sur la deuxième syllabe (an**co**ra). La langue arabe appartient à cette deuxième catégorie (cf. *infra*).

En plus de l'accent lexical, considéré comme l'accent principal dans le mot, il existe dans certaines langues un deuxième type d'accent qualifié d'*accent secondaire*. Celui-ci se présente sur les mots polysyllabiques et a une fonction différente de son analogue primaire.



En français par exemple, sa fonction selon Padeloup est la régulation rythmique de l'énoncé [Pas90]. En arabe, l'accent secondaire n'a pas de rôle très déterminé au niveau de la perception. L'accent porté par les syllabes non accentuées est appelé *accent de troisième niveau* ou « *wakstress* ».

Par ailleurs, il existe une différence entre l'accent lexical et l'accent *emphatique*. Alors que le premier intègre le mot dans une prononciation « neutre » et dont les règles de placement sont inhérentes à la langue, le second est introduit volontairement par le locuteur pour la mise en relief d'un mot dans son contexte. Par exemple : « **tu** dois partir » (c'est toi et pas quelqu'un d'autre) ; « tu **dois** partir » (marquer l'obligation de partir).

Les mots d'une langue n'ont pas toujours le même statut vis-à-vis de l'accent lexical. Ils sont généralement classés en deux catégories : *les clitiques*, qui ne reçoivent pas d'accent de mot, et les *non-clitiques* ou *mots accentogènes* qui reçoivent cet accent. Ce sont les catégories grammaticales des mots qui déterminent leur nature clitique/non clitique : les mots grammaticaux (particules monosyllabiques de l'arabe, prépositions monosyllabiques...) sont des clitiques, et les mots lexicaux des non-clitiques. Cependant, un clitique peut recevoir un accent dans certains cas. Exemple : *في المنزل* /filmanzili/ (« dans la maison ») (*في* /fi/ est un clitique) ; *نمت فيه* /nimtu flhi/ (« j'ai dormi à l'intérieur ») (*فيه* est un non-clitique). Ici, la préposition monosyllabique *في* à laquelle est attaché un suffixe *o* porte un accent lexical.

### 6.2.2. Étude de l'accent arabe

L'étude de l'accent en langue arabe standard a longtemps été occultée par les grammairiens, et ce pour plusieurs raisons. D'abord, la diversité et l'influence des dialectes arabes ne permettaient pas l'ébauche d'une étude uniformisée qui soit admise par tous. Ensuite, le rôle de l'accent n'est pas évident au premier abord, au point que certains linguistes ont nié son existence [Raj89]. Leur argumentation était que le placement de l'accent, s'il existait, sur n'importe quelle syllabe du mot n'affecte aucunement le sens de celui-ci.

Ces dernières années, nous avons noté un certain engouement pour l'étude de la prosodie arabe, ce qui a consolidé l'hypothèse selon laquelle l'accent lexical existe en arabe [Elg01] [Han00]. Bohas [Boh79] quant à lui, admet non seulement son existence, mais il cite des exemples où celui-ci joue un rôle distinctif. Ce postulat a été ensuite confirmé par Rajouani [Raj89] dans son article « Sur l'étude de l'accent lexical ». Ainsi, ce dernier a établi par des tests perceptifs que le mot *فَعَلَا* /faʕalA/ a le sens de « ils ont fait » (au duel) si l'accent est porté par la première syllabe *فَ* /fa/, et « ils ont pris de l'altitude » s'il est porté par la deuxième syllabe *عَ* /ʕa/. Notons que dans ce deuxième cas, la transcription orthographique du mot est *فَعَلَى* /faʕalA/, avec la syllabe *فَ* /fa/ comme préfixe.

Mais les exemples peu nombreux dans la littérature sont-ils suffisants pour justifier de l'importance de l'accent lexical arabe? Nous pensons pour notre part qu'au-delà de sa fonction linguistique, la modélisation de l'accent lexical est un élément **capital** pour la génération d'une voix synthétique qui se veut proche de la voix naturelle.

#### A. Système syllabique de la langue arabe

La langue arabe comporte cinq types de syllabes classées selon les traits ouvert/fermé et court/long (C=consonne, V=voyelle) (cf. figure 15). Une syllabe est dite *ouverte* (resp. *fermée*) si elle se termine par une voyelle (resp. une consonne). Elle est dite *courte* si elle est ouverte et intègre une voyelle brève. Elle est dite *longue* si elle se termine par une voyelle longue ou par une consonne. La terminologie *syllabe lourde* (CVV, CVC) et *syllabe sur-lourde* (CVVC, CVCC) est également utilisée dans la littérature. La syllabe CVVCC n'est pas attestée en arabe standard.

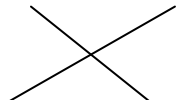
	<b>courte</b>	<b>Longue</b>
<b>ouverte</b>	CV	CVV
<b>fermée</b>		CVC, CVVC, CVCC

Fig. 15 : Système syllabique de la langue arabe.

Le système syllabique de l'arabe a les caractéristiques suivantes [Ela70] :

- Toutes les syllabes commencent par une consonne suivie d'une voyelle ;
- Les syllabes comportent une seule voyelle ;
- La syllabe CV peut se trouver en début, en milieu ou en fin de mot. Exemple : دَرَبَ/Daraba/ CVCVCV (« il a frappé ») ;
- La syllabe CVV peut se trouver en début, en milieu, en fin de mot ou isolée. Exemples : كَانَ /kAna/ CVVCV (« il était »), بَاقِي /baAqI/ CVVVCV (« le reste »), فِي /fI/ CVV (« dans ») ;
- La syllabe CVC peut se trouver en début, en milieu, en fin de mot ou isolée. Exemples : بَعْدَ /baeda/ CVCCV (« après »), مُدْ /mud/ CVC (« donne »),
- La syllabe CVVC peut se trouver en début, en milieu, en fin de mot ou isolée. Exemples : كَبِيرٌ /kabIr/ CVCVVC (« grand ») ;

- La syllabe CVCC se trouve uniquement en fin de mot ou isolée. Exemple : مَهْرُ /mahr/ CVCC (« perle ») ;

## B. Position de l'accent lexical

La langue arabe est une langue à *accent variable*. Tous les auteurs admettent que celui-ci est prédictible mais ils diffèrent quant aux règles régissant sa place dans le mot. Pour Cantineau [Can60], la position de l'accent est subjective : elle dépend de paramètres liés directement au locuteur (dialecte, culture, etc.).

Selon Kouloughli [Kou76], l'accent est limité aux trois dernières syllabes du mot. Il propose les règles suivantes sur la position de l'accent dans le mot :

- Si la dernière syllabe du mot est une sur-lourde, alors elle porte l'accent lexical.
- Si la règle précédente ne s'applique pas et si la pénultième est une syllabe lourde, alors elle porte l'accent lexical.
- Si les deux règles précédentes ne s'appliquent alors l'antépénultième porte l'accent lexical.

Ainsi, selon l'auteur, l'accent ne remonte jamais au-delà de l'antépénultième. Le point de vue de El-Ani [Ela70] est différent à ce sujet. Celui-ci considère l'existence de trois niveaux d'accents : l'accent primaire, l'accent secondaire et l'accent de troisième niveau (niveau inaccentué). La position de ces accents dépend de la structure syllabique du mot : les monosyllabiques portent un accent primaire et un accent de troisième niveau, alors que les polysyllabiques portent un accent primaire, un accent secondaire et un accent de troisième niveau. El-Ani propose les règles suivantes sur la position de l'accent dans le mot :

- Si le mot est constitué uniquement de syllabes de type CV, la première syllabe porte alors l'accent primaire et les autres syllabes l'accent de troisième niveau.
- Si le mot contient une seule syllabe longue, elle porte alors l'accent primaire et les autres syllabes l'accent de troisième niveau. Les syllabes longues en fin de mot sont ignorées.
- Si le mot est constitué de deux syllabes longues ou plus, la syllabe longue la plus proche de la fin du mot porte alors l'accent primaire, la syllabe longue la plus proche du début du mot porte l'accent secondaire et les autres syllabes l'accent de troisième niveau. Les syllabes longues en fin de mot sont ignorées.

Ces règles dénotent que l'accent peut se produire sur toutes les syllabes du mot, à l'exception de la dernière syllabe. Elles ne distinguent pas syllabe lourde et syllabe sur-lourde et font remonter l'accent au-delà de l'antépénultième.

Plusieurs travaux se sont intéressés à la validation des règles sur la position de l'accent lexical du point de vue des variations de la fréquence fondamentale. Celle-ci constitue le paramètre le plus pertinent pour la perception de cet accent (cf. chapitre 8). Parmi ces travaux, ceux de Es-Kali [Esk88], El-Kafi [Kaf90], Hanna [Han00] et Safa [Saf01] qui ont adopté les règles de El-Ani dans leurs travaux.

### 6.2.3. Conclusion

Nous avons présenté dans cette section les caractéristiques syllabiques de la langue arabe et quelques points de vue sur la position de l'accent lexical. Par rapport à ce dernier point, nous restons prudents sur le caractère général des règles proposées dans la littérature en raison des influences dialectales, ce qui peut expliquer les divergences d'opinion. En cela, nous rejoignons la position de Cantineau [Can60].

Etant donné que notre objectif est de développer un modèle prosodique pour la SAT, nous avons choisi de travailler sur une voix unique afin d'obtenir une description cohérente, basée sur des variations individuelles plutôt que sur une série d'invariants pour la langue traitée. En effet, les caractéristiques individuelles d'une voix favorisent la synthèse de parole moins mécanique, moins monotone et proche du naturel [Kel92]. En retour, la synthèse de la parole permet une validation perceptive de certaines hypothèses. Ainsi, nous proposerons des règles de placement de l'accent lexical en prenant en compte les résultats d'analyse de la voix en question, sans prétendre à la généralité des règles proposées.

## **PARTIE 3 : PROSODIE**

## Chapitre 7 : Etude de la prosodie

### 7.1. Généralités

Le terme *prosodie* recèle des notions différentes selon le point de vue adopté pour son étude. Du point de vue acoustique, la prosodie se définit au moyen des paramètres de la fréquence du fondamental (estimation du son laryngien à un instant donné sur le signal), de la durée (intervalle de temps entre deux points sur le signal) et de l'intensité (énergie contenue dans le signal). Du point de vue de la perception de la parole, elle concerne l'étude des phénomènes de *l'accentuation* et de *l'intonation* (variation de hauteur, de rythme et d'intensité) permettant de véhiculer de l'information liée au sens de la phrase [Boi00]. Toutefois, il est difficile d'établir une correspondance directe entre paramètres physiques et corrélats perceptifs.

Le terme *suprasegmental* est également employé dans la littérature pour désigner la prosodie. Il indique que celle-ci relève de phénomènes qui dépassent le cadre du phonème, s'étendant de la syllabe jusqu'à la phrase, atteignant pour certains le paragraphe [Van99a]. Ainsi, l'accent, l'intonation et le rythme sont des phénomènes suprasegmentaux. Par opposition, les phénomènes dont la portée ne franchit pas les frontières du phonème sont qualifiés de phénomènes *segmentaux*.

Dans beaucoup de langues, la prosodie présente des *traits* similaires qui pourraient être *universels*. La tendance de l'intonation à descendre à la fin des phrases assertives et à augmenter à la fin des phrases interrogatives constitue sans doute l'exemple le plus significatif. Cependant, cette universalité ne doit pas occulter les spécificités inhérentes à chaque langue où à chaque individu dans la réalisation prosodique des énoncés : un texte peut être prononcé de manières différentes selon les caractéristiques anatomiques du locuteur, son origine régionale, sociale, son état émotif, son tempérament, etc.

Plusieurs recherches soulignent que les niveaux linguistiques entretiennent une relation privilégiée avec la prosodie en ce qui concerne les langues indo-européennes. Au niveau syntaxique, il existe plusieurs réalisations prosodiques possibles pour une structure syntaxique donnée, mais pas une infinité ou n'importe lesquelles. À l'inverse, une réalisation prosodique donnée ne correspond qu'à quelques structures syntaxiques particulières. Mertens [Mer00] envisage trois configurations possibles liant la syntaxe à la prosodie : soit la syntaxe et la prosodie sont complètement indépendantes, soit elles sont le reflet exact l'une de l'autre, ou encore elles entretiennent des relations privilégiées, dites de **congruence** (de non contradiction). Au niveau *sémantique*, la réalisation prosodique d'un énoncé (modalité, placement des pauses, etc.) révèle des informations qui se rapportent directement au *sens* de celui-ci.

De ce fait, le calcul des structures prosodiques en SAT passe par la prise en compte des connaissances linguistiques, essentiellement de nature syntaxique et rythmique. Pour une phrase donnée, le problème est alors de trouver la réalisation prosodique la plus *acceptable* parmi l'ensemble des réalisations possibles, au regard des contraintes exprimées par ces connaissances.

## **7.2. Fonction de la prosodie**

La prosodie a plusieurs fonctions qui peuvent se classer en deux catégories : linguistiques, véhiculant des informations modales et structurales sur l'énoncé (expression de la modalité, segmentation du continuum et hiérarchisation) et para-linguistiques, sur le locuteur et ses rapports avec son discours et ses interlocuteurs.

- La fonction *expression de la modalité* permet à l'auditeur d'identifier le mode de la phrase de son interlocuteur (assertion, question, exclamation, ordre). Ainsi, une intonation montante sur la frontière des unités prosodiques traduit la continuation, alors qu'une intonation descendante traduit une finalité. Aussi, une intonation haute utilisée dans les questions oui/non manifeste l'attente d'une réponse.
- La fonction *segmentation* offre à l'auditeur des repères dans le continuum qui lui permettent de distinguer et d'incorporer mentalement les unités qui le composent : les débuts et fins de paragraphes, les débuts et fins de phrases et les frontières de groupes syntaxiques intègrent des ruptures qui sont plus ou moins marquées. À l'inverse, l'absence de rupture permet de préserver la relation syntaxique entre deux éléments contigus.
- La fonction *hiérarchisation* permet de dévoiler la structuration de l'énoncé en plusieurs niveaux (dérivation, emboîtement des constituants), en séparant les informations de premier plan, de second plan, etc. Pour Lacheret (chapitre « synthèse de la parole » dans [Sch02]), les énoncés possèdent des marqueurs intonatifs hiérarchisant qui rendent compte de ces niveaux de dérivation et d'imbrication syntaxique.
- Enfin, la prosodie permet de transmettre des informations de nature para-linguistique sur l'état émotionnel du locuteur (triste, ennuyé, enthousiaste, etc.), son appartenance géographique (accent régional), son âge, son sexe, etc. Elle fournit également des informations sur son attitude envers son discours (degré de son adhésion à son énoncé), et sur la gestion du dialogue face à ses interlocuteurs (hésitation, assurance, etc.).

### 7.3. La substance phonétique de la prosodie

Au niveau acoustique, l'accentuation s'accompagne de variations en proportion différente des trois paramètres de la prosodie (fréquence fondamentale, durée et intensité). En ce qui concerne les langues européennes les plus répandues, l'accent lexical est encodé avant tout par la fréquence fondamentale, alors que les paramètres de l'amplitude et de la durée ne jouent que des rôles de soutien [Kel92]. La hiérarchie admise comme étant universelle de ces trois paramètres, quant à leur importance, est : la fréquence fondamentale, la durée et l'intensité. Seuls les deux premiers paramètres sont pris en compte pour la génération des langues indo-européennes dans le système de synthèse d'Elan Speech.

Les recherches en prosodie de l'arabe ont confirmé l'affirmation selon laquelle la fréquence fondamentale est le paramètre le plus pertinent dans la perception de l'accent lexical, mais reste assez mitigé à propos du rôle de la durée et de l'intensité. Es-Skali [Esk88] a mené une étude instrumentale sur la pertinence de la fréquence fondamentale et de l'intensité et est arrivé aux conclusions suivantes :

- Le fondamental d'une syllabe accentuée se situe en moyenne entre 6 et 9 unités de perception au-dessus de celui d'une syllabe non accentuée (une unité de perception correspond à 5% de la valeur du fondamental d'une syllabe non accentuée) ;
- L'intensité d'une syllabe accentuée se situe en moyenne entre 4 et 6 dB au-dessus de celle d'une syllabe non accentuée. Cette différence n'est pas perçue à l'audition car elle ne dépasse pas le seuil de perception de 8 dB défini par Rossi [Ros71].

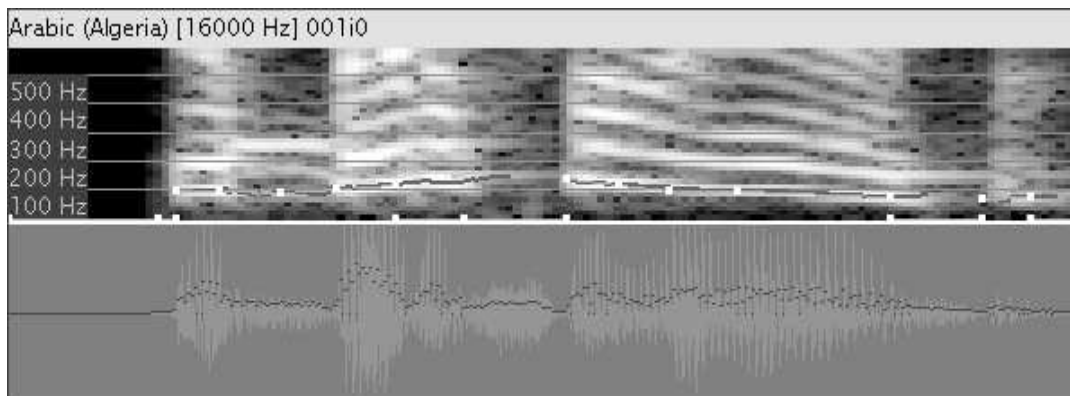
Une autre expérience menée par Rajouani [Raj89] a montré que l'intensité est moins significative que le fondamental, alors que la durée est moins importante que l'intensité. Ainsi, il a suggéré que la hiérarchie des 3 paramètres concourant à la perception de l'accent lexical soit dans l'ordre suivant : fréquence fondamentale, intensité et durée, ce qui n'est pas en adéquation avec la hiérarchie habituelle.

#### 7.3.1. La fréquence fondamentale

La fréquence fondamentale ou fréquence laryngienne (notée F0) représente la fréquence de vibration des cordes vocales. Son estimation est liée à la localisation de portions voisées sur le signal de parole, les sons non voisés ayant une fréquence nulle. Les algorithmes d'extraction de F0 peuvent être de type *temporel* (AMDF, LPC...) ou *fréquentiel*. Les premiers se basent directement sur la description temporelle du signal pour le calcul de F0 ( $F0=1/\text{Période}$ ), alors que les seconds s'appuient sur les fréquences des *harmoniques* (fréquence de résonance) qui peuvent être représentés graphiquement sur un spectrogramme.



La figure 16 représente une description spectrale (fenêtre du haut) et temporelle (fenêtre du bas) de la phrase *الدّرس العاشر* /ʔaddarsul ɛAOir/ (« le dixième cours »). Les trames blanches correspondant aux harmoniques sur le spectrogramme désignent les portions voisées (voyelles /a/ /u/ /i/) du signal acoustique.



**Fig. 16 : Représentations fréquentielle et temporelle de la phrase /ʔaddarsul ɛAOir/.**

Plusieurs algorithmes automatiques et semi-automatiques ont été réalisés pour l'extraction de F0, ce qui est d'un grand intérêt pour l'étiquetage prosodique compte tenu des coûts manuels d'une telle opération. Parmi eux, l'algorithme MOMEL (MODélisation MELodique) qui a été employé par Campione [Cam01] pour la stylisation de la courbe de F0 en vue de la transcription prosodique de grands corpus oraux. Cet algorithme réduit la courbe originelle en une suite de *points cibles*, puis les relie en utilisant une courbe *spline quadratique*. Cet algorithme est indépendant de la langue traitée et ne nécessite aucune pré-segmentation du signal acoustique. Une évaluation de MOMEL a montré qu'il produisait environ 5% d'erreurs sur les points cibles.

Dans ce travail, l'extraction du pitch est réalisée à l'aide de *Prosel*, un outil de recopie de prosodie naturelle développé par Elan Speech [Bou01]. À l'inverse de MOMEL, cet outil opère une segmentation du signal naturel avant le calcul du pitch selon le principe suivant (cf. figure 17) : *Prosel* extrait des paramètres acoustiques à partir des signaux naturel et synthétique correspondant au texte d'entrée, puis aligne ces deux signaux en utilisant l'algorithme DTW (Dynamic Time Wrapping). Les phonèmes de la voix synthétique sont ainsi alignés avec des zones de signaux naturels selon des critères de ressemblance basés sur les paramètres calculés (excluant le pitch), ce qui permet de déduire les limites temporelles des phonèmes sur la voix naturelle.

Le pitch est ensuite extrait sur les portions voisées de la voix naturelle en calculant d'abord trois valeurs par trame, puis en retenant celle qui offre la courbe la plus cohérente. Ainsi, la segmentation préalable du signal contribue à réduire les erreurs de calcul du pitch directement sur l'ensemble du signal.

La courbe mélodique est ensuite simplifiée en reliant les différents points par des segments de droite : c'est le principe de la stylisation qui se fonde sur l'hypothèse selon laquelle un certain nombre d'événements présents sur le signal acoustique peut être éliminé sans occasionner de changement au niveau de la perception [Bea94].

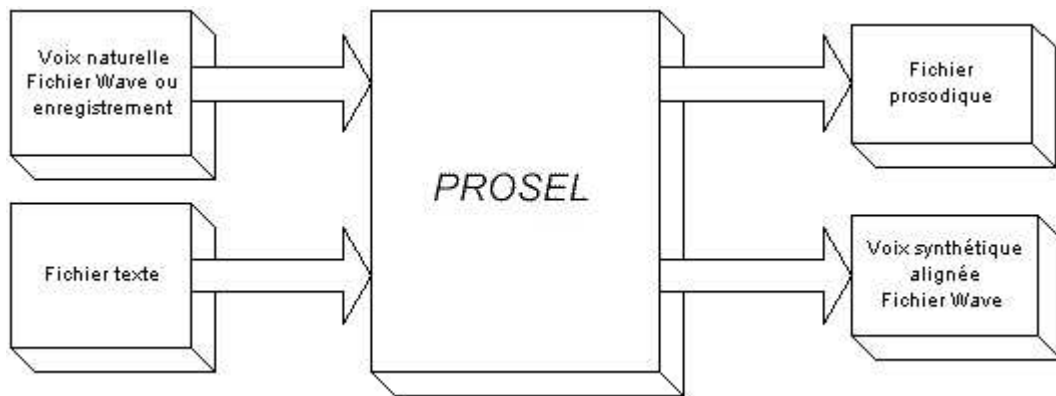


Fig. 17 : Vue générale de Prosel.

Sur le plan de la perception, la valeur de F0 correspond en première approximation à la sensation de *hauteur* que procure un son. Les phonéticiens utilisent le *ton* pour exprimer le rapport de hauteur entre une fréquence F1 et une fréquence F2, étant donné que notre oreille a une perception logarithmique de la hauteur et non pas linéaire. Ce rapport est calculé selon la formule suivante :

$$F = 6 \times ech \times \frac{\ln\left(\frac{F_2}{F_1}\right)}{\ln 2}$$

Dans cette formule, si  $ech=2$  alors F est exprimé en demi-ton, si  $ech=4$ , F est exprimé en quart de ton, etc.

### 7.3.2. La durée

Pour beaucoup d'auteurs, le paramètre de la durée est difficile à calculer car il ne dépend d'aucun *corrélat biologique*, contrairement à F0 et à l'intensité (qui dépendent respectivement de la tension des cordes vocales et de la pression sous-glottique). Pour calculer la durée d'un phénomène, il faudrait se fixer deux événements qui délimitent ses repères initial et final. Sur un signal de parole, cette tâche incombe au processus de *segmentation* qui, pour beaucoup de systèmes actuels, est basé sur le phonème.

Pour Klatt [Kla76], la durée est corrélée à une multitude de facteurs complexes de nature linguistique (accent, position des mots dans la phrase, catégorie grammaticale, etc.) et extra-linguistique (débit de parole, expressivité, etc.). Certains d'entre eux peuvent être

privilégiés par rapport à d'autres selon le type de corpus d'analyse et le style de lecture employés :

- L'analyse de **logatomes** dans des phrases porteuses, dont le but est d'uniformiser l'environnement linguistique (syntaxique, sémantique...), ne rend compte que des phénomènes intra-mots (phonétique, phonologique, etc.) ;
- L'analyse de **phrases** rend compte de phénomènes d'interaction entre les mots (syntaxique, sémantique, etc.). La difficulté ici est de faire la part des choses entre ce qui relève du contexte phonétique, syntaxique, sémantique, etc. De plus, les proportions d'influence des facteurs en question peuvent être de degrés divers.
- L'analyse de **corpus de lecture spontanée** rend compte de phénomènes liés à l'hésitation, etc.
- Enfin, plusieurs **lectures de locuteurs différents** rendent compte des variabilités individuelles (physiologique, régionale, etc.) par rapport aux autres variabilités.

#### Modèles de prédiction de la durée

Barbosa [Bar94] puis Morlec [Mor97] ont classé les modèles de prédiction de la durée selon la taille du segment qui fait l'objet du calcul. Celui-ci peut être le phonème, la syllabe ou des unités de taille plus grande comme le **P-center**. Nous reprenons ce classement dans ce qui suit.

- Le phonème

Les modèles qui prédisent directement la durée phonémique peuvent être classés en trois catégories : additive, multiplicative ou additive/multiplicative. Dans chacun des cas, une durée intrinsèque est calculée pour chaque phonème, puis déformée par ajout/soustraction ou multiplication/division d'une *valeur* qui caractérise les facteurs contextuels agissant sur le phonème. Parmi ces modèles, nous citons ceux de Klatt [Kla76], d'O'Shaughnessy [Osh81] pour les langues européennes, et les modèles de Zemirli [Zem00], de Amrouche [Amr98] et de Ghazali [Gha92a] pour la prédiction des voyelles de la langue arabe.

Le modèle multiplicatif de France Télécom R&D [Bar87], implémenté dans un système de SAT du français basé sur la concaténation de diphtonges, modifie la valeur intrinsèque des phonèmes selon plusieurs critères (position dans le mot, contexte phonétique, catégorie grammaticale, structure syntaxique). Les formules utilisées sont les suivantes :

$$\text{Durée\_voyelle} = DI.Vi.mc$$

$$\text{Durée\_consonne} = \text{DI.Cij}$$

Où  $D_i$  représente la durée intrinsèque du phonème,  $V_i$  et  $C_{ij}$  sont respectivement des coefficients qui rendent compte de l'influence du contexte pour les voyelles et les consonnes et  $m_c$  des coefficients co-intrinsèques dans le cas des voyelles.

- La syllabe

Parmi les modèles qui prédisent les durées de syllabe, décrivons celui de Campbell [Cam92] pour l'anglais britannique. Ce modèle procède en deux étapes :

1. Le calcul de la durée de chaque syllabe sur la base de connaissances phonologiques et phonotactiques. Ces durées syllabiques sont préalablement mesurées en utilisant une approche par apprentissage.
2. Le calcul de la durée de chaque phonème par *répartition* des durées syllabiques. Campbell fait l'hypothèse que les phonèmes à l'intérieur d'une syllabe subissent tous la même déformation : un coefficient de *z-score* (ou variable centrée) est extrait pour chaque phonème à partir du corpus d'analyse en utilisant la formule suivante :

$$z = \frac{\text{durée\_observée\_du\_phonème} - \text{moyenne\_du\_phonème}}{\text{écart\_type\_du\_phonème}}$$

La durée syllabique représente la somme des durées de chaque phonème. Elle est calculée à l'aide de l'équation suivante :

$$\text{durée}(\text{syllabe\_i}) = \sum_{i=1}^{\text{nb\_phonèmes}} \exp(\text{MoyennePhonème\_i} + z\_score * \text{EcartypePhonème\_i})$$

- Le GIPC (Groupe-Inter-Perceptuel-Center)

Les structures rythmiques de la parole sont caractérisées par le retour de *formes sonores* au niveau **perceptif** à des périodes quasi-régulières. Une distinction est alors faite entre les langues à *chronométrage syllabique* et les langues à *chronométrage accentuel* : dans les premières, les syllabes sont perçues à des intervalles réguliers (comme l'anglais) ; dans les secondes, les accents de mots sont perçus à des intervalles réguliers (comme le français).

Le Perceptual-Center est défini comme le point de repère de perception dans une chaîne sonore. La distance entre deux P-Center successifs est donc identique dans les langues à chronométrage syllabique et accentuel.

Barbosa [Bar94] a présenté un modèle qui se base sur le GIPC, une unité qui s'étale entre deux P-Center successifs, où chaque P-Center coïncide avec le début d'une voyelle.

Pour lui, cette unité est une alternative à la syllabe pour la structuration du rythme. Il utilise ensuite le modèle de répartition de Campbell pour le calcul de la durée des phonèmes à l'intérieur du GIPC.

### 7.3.3. L'intensité

Résultant de la pression sous-glottique, l'intensité est le paramètre prosodique le plus simple à calculer. Elle est mesurée sur des portions de signal allant de 5 à 10 ms (énergie à court terme) et exprimée en décibels pour respecter l'échelle perceptive. Sa formule est la suivante :

$$E_{ab} = 10 \times \log_{10} \left( \sum_{t=A}^T S_t^2 \right)$$

Ce paramètre est le plus souvent négligé en génération de la prosodie. Bien que considéré comme dépendant de F0, ces deux paramètres peuvent varier indépendamment dans l'interrogation [Kel92].

En ce qui nous concerne, nous nous basons sur le paramètre de F0 pour la mise en relief de la syllabe accentuée sans pour autant nous prononcer sur le rôle que pourrait jouer l'intensité. L'étude de ce paramètre fera l'objet de nos travaux futurs.

## 7.4. L'intonation

Le terme *intonation* a deux définitions possibles : au sens strict, ce mot désigne les changements relatifs à la hauteur de la voix, que certains chercheurs confondent avec le mot *mélodie*. Le sens le plus étendu de ce terme fait aussi référence aux changements de la durée et de l'intensité. Dans ce dernier cas, il s'apparente au mot *prosodie*. Nous utiliserons pour la suite de ce document le terme *intonation* en référence aux variations de la hauteur, dont le corrélat acoustique est le paramètre F0.

L'étendue de l'intonation est la phrase. C'est un phénomène *macro-mélodique* de portée globale, par opposition aux phénomènes locaux dont la portée ne dépasse pas les limites du mot (accent de mot par exemple). Des pics intonatifs peuvent apparaître sur la courbe de l'intonation, correspondant à des phénomènes dits *micro-mélodiques* : « *Ce sont des phénomènes qui résultent de contraintes physiologiques et/ou acoustiques sur l'appareil phonatoire. Ils n'ont aucune incidence sur la perception globale de l'intonation* » [Bea94].

### 7.4.1. La déclinaison

Le phénomène de déclinaison représente la tendance de la fréquence fondamentale à décroître du début à la fin de la phrase. Il se manifeste dans un certain nombre de langues

(français, anglais, italien, etc.), notamment dans la langue arabe [Bal02b] [Zak00a] et pourrait bien être universel [Vai95]. La figure 18 représente l'évolution de F0 pour la phrase وَقَعَ نَظْرُهُ عَلَى نَمْلَةٍ /waqaea naZruhu calA namlatin/ (« il a aperçu une fourmi »).

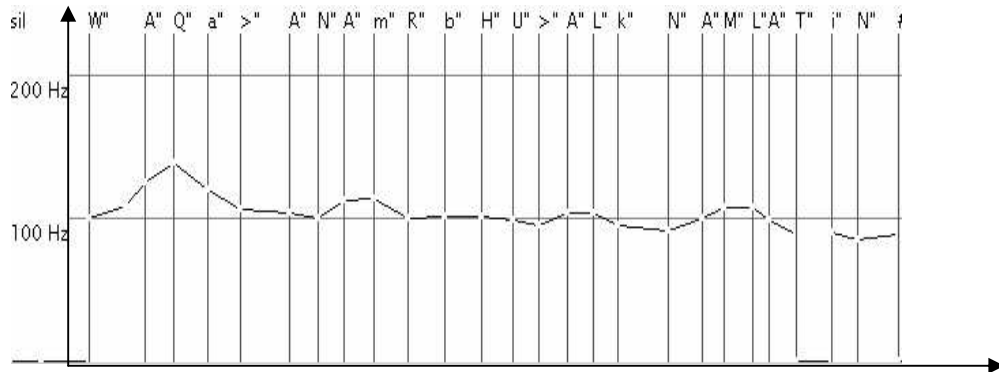


Fig. 18 : Evolution de F0 pour la phrase /waqaea naZruhu calA namlatin/.

Pour décrire la déclinaison, plusieurs auteurs ont établi une ligne sur laquelle viennent se superposer des événements locaux (maxima et minima). En général, le taux de déclinaison de cette ligne est fonction de la longueur de la phrase qu'elle représente : plus la phrase est courte, plus la pente de la déclinaison est importante. Pour l'estimation de cette pente, beaucoup de travaux ont opté pour la méthode des moindres carrés qui calcule la meilleure droite passant par des valeurs de F0 prélevées sur le signal [Van99a].

Beaugendre [Bea94] a défini deux lignes de déclinaison (une ligne haute et une ligne basse) qui délimitent *le registre du locuteur*, c'est-à-dire, la bande à l'intérieur de laquelle se propage la fréquence du fondamental (cf. figure 19). Pour Safa [Saf01], la ligne haute relie la valeur de F0 sur la première syllabe accentuée avec la valeur de F0 sur la dernière syllabe accentuée ; la ligne basse relie la valeur de F0 sur la dernière syllabe du premier mot avec la valeur de F0 de la dernière syllabe de la phrase, quand celle-ci est de type assertive.

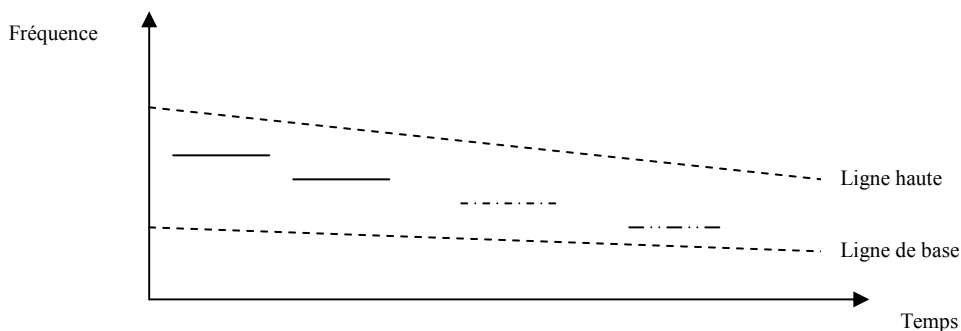
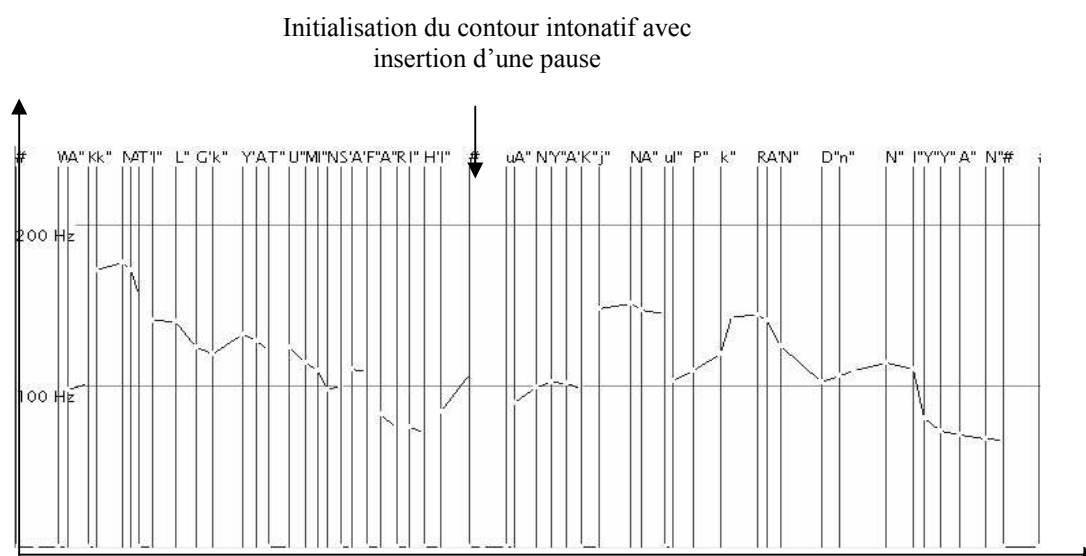


Fig. 19 : Représentation du registre du locuteur.

Les phrases longues sont souvent accompagnées d'une réinitialisation de la courbe intonative qui intervient généralement aux frontières de groupes de souffle [Boi00]. Ces derniers correspondent à des segments à l'intérieur desquels aucune pause n'est marquée. Cependant, il arrive que cette réinitialisation se produise sans une marque de pause, de même qu'une pause peut avoir lieu sans une réinitialisation de la courbe mélodique.

La difficulté en modélisation de la réinitialisation est de savoir quand elle doit intervenir dans la phrase. Souvent, elle coïncide avec une frontière syntaxique compte tenu du temps (ou du nombre de syllabes) écoulé depuis la dernière remise à zéro de la courbe intonative. Dans notre modèle, la réinitialisation intervient à la frontière d'un tronçon, en même temps qu'une pause, si le nombre de syllabes depuis la dernière pause et le nombre de syllabe jusqu'au point de ponctuation suivant sont suffisants (cf. chapitre 8) [Bal02b].



**Fig. 20 : Exemple de réinitialisation de la courbe intonative.**

La figure 20 représente un exemple de réinitialisation de la courbe mélodique dans la phrase *وَكَانَتِ الْغَايَةُ مِنْ سَفَرِهِ أَنْ يَكُونَ إِيطَارًا دِينِيًّا* /wa kAnatil GAYatu min safarihi ?an yakUna ?ITArAn dIniyyan/ («l'objectif de sa visite était d'offrir un cadre spirituel»). Cette réinitialisation intervient à la frontière initiale du groupe syntaxique verbal /?an yakUna/ après 12 syllabes à partir du début de la phrase.

#### 7.4.2. Modèle de génération de l'intonation

Les modèles actuels de génération de F0 peuvent se classer en trois grandes catégories selon qu'ils s'appuient sur une approche physiologique (commande mélodique), phonologique (points cibles, tons), sur la concaténation de contours intonatifs stockés ou par

apprentissage. Ces modèles trouvent dans la synthèse de la parole un cadre de test et de validation très intéressant pour les différentes langues étudiées.

#### A. Modèles de commande mélodique

Ces modèles tentent de reproduire les mécanismes *physiologiques* qui gouvernent la production de F0. Le premier modèle de ce type a été proposé par Ohman [Ohm67] qui considérait que la courbe intonative résultait de la *superposition* de deux types de contours : le premier est lié à une commande de *phrase* (il correspond à une réponse d'un système linéaire de second ordre à des impulsions) ; le second est lié à une commande *d'accent* (il correspond à la réponse d'un autre système linéaire de second ordre à des fonctions échelons). Ces travaux ont inspiré le modèle de Fujisaki [Fuj67].

#### B. Approches phonologiques

Cette approche utilise une démarche *ascendante*. Elle consiste à extraire de la substance phonétique une représentation simplifiée de la courbe mélodique et à la relier à un modèle *phonologique*. Elle compte deux méthodes de modélisation : la méthode par point cible et l'école hollandaise.

##### **La méthode par point cible**

Dans cette méthode, la courbe mélodique est réduite en une séquence de points sur le plan temps/fréquence, appelés *points cibles*. Ces points correspondent généralement aux *extremas* de cette courbe et sont jugés **suffisants** pour transporter l'essentiel de l'information mélodique [Pie81]. Les transitions entre les points cibles sont considérées comme *non pertinentes* au niveau de la perception. Pour relier ces points, les auteurs ont choisi des fonctions de transition quadratiques [Pie81] ou linéaires [Bea94], en estimant que ce choix n'a pas d'impact sur la perception globale de la mélodie.

Des symboles sont ensuite associés aux différents points cibles, ce qui permet de dériver une *représentation formelle* à partir de la substance phonétique. Pour la description de l'anglais américain, Pierrehumbert utilise deux catégories de *tons relatifs*, haut (H) et bas (L), dont la séquence est régie par une grammaire d'états finis. Elle distingue un ton d'accent de pitch, un ton de frontière initiale, un ton d'accent de phrase et un ton de frontière finale. La grammaire appliquée permet la description de la phrase en plusieurs niveaux : niveau *mot prosodique* délimité par un ton d'accent de pitch, *niveau phrase intermédiaire* formé de mots prosodiques et niveau *phrase intonative* constitué de phrases intermédiaires.

##### **L'école hollandaise**

La courbe mélodique est simplifiée en une suite de segments de droites sur la base de la *tolérance perceptive*, l'objectif étant que la courbe obtenue soit **indiscernable** de l'originale [Bea94]. Ainsi, après la stylisation de la courbe mélodique, Beaugendre classe les



mouvements obtenus dans des classes standards, puis les aligne avec les structures linguistiques pour déterminer ceux qui modélisent l'accent lexical, les frontières prosodiques montantes ou descendantes, etc. Ce sont ces mouvements *standardisés* qui seront utilisés pour former les contours synthétiques. Leurs agencements sont spécifiés dans une grammaire déterministe sur la base de contraintes structurelles et linguistiques.

### C. Modèles par contours stockés

Sur le même principe que les méthodes de synthèse de la parole basées sur la concaténation d'unités acoustiques (diphones, di-syllabes, etc.), cette approche vise à construire le contour mélodique global de la phrase en mettant bout à bout des contours de signaux élémentaires, de taille variable, prélevés directement sur un corpus de parole. Parmi les travaux qui utilisent cette démarche, citons celui d'Aubergé [Aub91].

Pour Aubergé, l'intonation et la syntaxe sont des structures indépendantes qui présenteraient des points de *rendez-vous*, établis par l'organisation syntaxique de l'énoncé. Ainsi, les contours mélodiques sont extraits et classés en tenant compte du type et de la longueur des groupes syntaxiques, où chaque classe reçoit un contour mélodique *type* correspondant à la moyenne des contours mélodiques qu'elle contient.

Les modèles exposés ci-dessus présentent l'avantage d'explicitement la relation qui lie la substance prosodique avec les niveaux linguistiques. Cependant, ils ne permettent pas de reproduire finement les caractéristiques de la courbe mélodique, en particulier les manifestations micro-prosodiques. L'apprentissage automatique permet de s'affranchir de cette limite.

### D. Modèles par apprentissage automatique

Cette méthode s'appuie sur l'utilisation de réseaux neuronaux qui associent description linguistique et substance prosodique. Ainsi, de la *pertinence* des paramètres linguistiques à l'entrée du réseau dépend la qualité du contour mélodique.

Les modèles de Sagisaka [Sag90] et Traber [Tra92] appartiennent à ce courant de travaux. Ce dernier favorise la description de l'accent lexical, le type de phonèmes et la valeur intrinsèque de F0 associée aux syllabes à l'entrée du réseau. Néanmoins, même si le résultat obtenu par ces méthodes est jugé satisfaisant, le principal inconvénient lié à leur utilisation est l'incapacité de savoir comment se fait le passage entre la description linguistique et les paramètres de sortie du réseau.

## 7.5. Etude de l'intonation arabe

Mrayati [Mra84] a proposé quatre modèles de contour intonatif arabe selon la modalité de la phrase (déclarative, interrogative, impérative, exclamative) (cf. figure 21). La fréquence fondamentale y est représentée sur une échelle à 4 niveaux, définie comme suit : /1/ basse, /2/ moyenne, /3/ haute et /4/ extra-haute. Toutefois, l'auteur n'a fourni aucune description physique concernant les échelles utilisées.

Modalité	Modèle de l'intonation
Déclarative	2-2-1
Impérative	2-3-1
Interrogative	3-2-1 ou 2-3-1
Exclamative	2-3-1

Fig. 21 : Modèles intonatifs des phrases arabes.

### 7.5.1 Le contour intonatif

Plusieurs recherches s'accordent à dire que l'accent lexical d'un mot est préservé au niveau de la phrase [Esk88] [Elk90]. La structure intonative d'une phrase est alors dérivée à partir de la description accentuelle des mots qui la composent. Le niveau des accents lexicaux (hauteur de F0) est corrélé à la modalité de la phrase d'une part et, selon les auteurs, à la structure syntaxique de l'énoncé [Raj89] ou à la position des syllabes accentuées dans la phrase d'autre part [Zak00a] [Bal02b].

Comparativement aux langues européennes, il existe peu de travaux sur la génération automatique de la prosodie arabe dans la littérature. Le MIR (Modèle Intonatif de Rabat) a néanmoins suscité quelque intérêt dans les années 90, compte tenu des publications apparues à son sujet. Développé à la Faculté des Sciences de Rabat (Maroc), il a inspiré les récents travaux sur la génération de la prosodie développés à l'ENSEIRB<sup>13</sup> [Saf01] [Zak00a]. Nous présentons dans ce qui suit les principales caractéristiques de MIR afin d'illustrer ses limites.

#### Présentation de MIR

La génération du contour intonatif d'une phrase est basée sur les faits suivants :

- Les pics mélodiques se situent sur les syllabes accentuées.
- La place de l'accent lexical est sauvegardée dans la phrase.

<sup>13</sup> Ecole Nationale Supérieure d'Electronique, d'Informatique et de Radiocommunications de Bordeaux

- Au niveau du mot, le contour mélodique se compose de deux segments linéaires : le premier montant, allant de la première syllabe FD jusqu'à la syllabe accentuée FA, et le second descendant allant de la syllabe accentuée FA jusqu'à la dernière syllabe FF (cf. figure 22, contour intonatif du mot  $\text{يَتَعَلَّم}$  /yataεallamu/ (« il apprend »)).

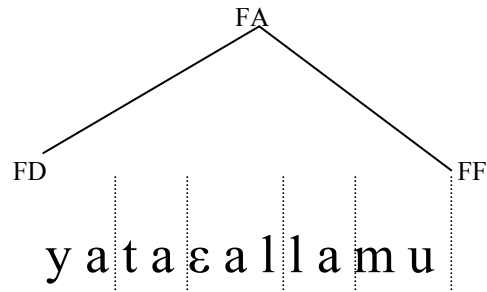


Fig. 22 : Contour intonatif au niveau du mot.

- Au niveau de la phrase, le contour mélodique est constitué par assemblage des contours de chaque mot (cf. figure 23, contour mélodique de la phrase  $\text{يُعَلِّمُ الْوَلَدَ}$  /yuεallimul walada/ -« il apprend à l'enfant »). Le maximum intonatif sur la syllabe FA dépend de la modalité de la phrase et de l'attribut syntaxique du groupe auquel il appartient.

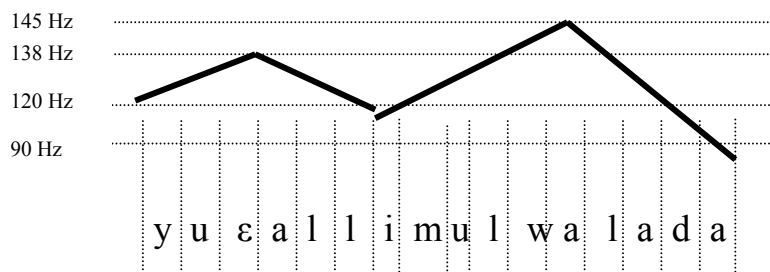


Fig. 23 : Exemple d'un contour intonatif pour une phrase en arabe.

- Les informations syntaxiques sont déterminées à partir de marqueurs intonatifs positionnés manuellement dans la phrase.

Les contours générés par MIR concernent des phrases de type déclaratif et interrogatif. Les caractéristiques physiques de ce modèle sont présentées dans le tableau 24.

		FD	FA	FF
<b>Phrase verbale</b>	Verbe	120	138	120 pour tous les mots sauf le dernier de la phrase pour lequel FF=100
	Chaque mot du groupe sujet	120	138	
	Chaque mot du groupe complément	120	145	
<b>Phrase nominale</b>	Chaque mot du groupe sujet	120	145	
	Chaque mot du groupe attribut	120	145	

**Fig. 24 : Règles du modèle intonatif MIR.**

### 7.5.2. Discussion

Il faut tout d'abord rappeler que notre démarche s'inscrit dans le cadre d'un système de SAT entièrement automatique. Il nous est donc inconcevable d'opter pour un étiquetage syntaxique manuel, à l'instar de MIR, afin d'effectuer par exemple la mise en relation des tronçons. Ensuite, le MIR ne s'applique qu'à des phrases à structure simple (groupe verbal + groupe nominal + groupe complément ou groupe sujet + groupe attribut), ce qui limite la couverture de ce modèle compte tenu de la diversité structurelle des phrases dans la pratique.

Une autre remarque est que le MIR ne rend pas compte du phénomène de déclinaison, alors qu'il semble y avoir un consensus ces dernières années au sujet de son existence en arabe [Naj98] (dans le MIR, les F0 sur les syllabes accentuées du groupe complément sont plus élevées que celles des syllabes accentuées du groupe verbal et du groupe sujet). De plus, la réinitialisation du contour intonatif n'est pas modélisée pour les phrases de taille importante.

L'autre modèle proposé dans la littérature [Saf01] ne fait aucunement référence à la syntaxe. La remise à zéro de F0 dans ce modèle se fonde uniquement sur des données phonotactiques, ne respectant pas ainsi une règle élémentaire en prosodie, selon laquelle la réinitialisation de la courbe intonative ne peut pas s'opérer à des endroits interdits, comme par exemple après la particule في /fi/ qui régit le cas indirect (à l'intérieur de groupes homogènes).

Un autre point sur lequel ces modèles ne se sont pas penchés est la *collision d'accent* : la succession de deux syllabes accentuées est interdite [Mar98]. Ainsi, dans la phrase في داره /fi dArihi/ (« dans sa maison »), l'accent normalement placé sur في /fi/ doit être supprimé du fait de la collision avec l'accent de /dArihi/. Enfin, les modèles présentés dans la littérature ne rendent pas compte de la place des pauses en génération de la prosodie.

## CHAPITRE 8 : Analyse et synthèse de la prosodie

Dans le système d'Elan Speech, la sortie de l'analyse syntaxique, qui fournit une séquence de mots étiquetés grammaticalement, alignée avec la séquence de tronçons, est connectée aux modules suivants de calcul de la prosodie. Une frontière mineure est pausée à la fin des tronçons (#fm), une frontière majeure est placée après un signe de ponctuation faible (#FM), une frontière terminale marque la fin de la phrase (#FT, qui peut être réalisée comme montante — interrogation — ou descendante). Exemple :

الغَنَاءُ (#fm) يُتِيحُ (#fm) لَنَا (#fm) التَّعْيِيرَ (#fm) عَنِ كَلِّ مَشَاعِرِنَا الْعَمِيقَةِ (#FT). فَيُفَضِّلُهُ  
(#fm) يُمَكِّنُنَا (#fm) أَنْ نُبَدِيَ (#fm) فَرْحَةَ غَامِرَةٍ (#FM)، أَوْ حُزْنَ عَمِيقًا (#FT).

Les unités délimitées par ces frontières ne constituent pas des groupes de souffle séparés par des pauses : l'ajustement avec le nombre de syllabes requiert un autre module qui a pour but de prendre en compte ces contraintes rythmiques. Après l'assignation de l'accent lexical, l'interface syntaxe-prosodie permet de générer les paramètres prosodiques de hauteur (cf. section 8.3) et de durée (cf. section 8.4).

### 8.1. Corpus d'analyse

Notre corpus d'analyse se compose de 168 phrases dont 148 sont déclaratives et 20 interrogatives, avec une moyenne de 9 mots par phrase. Ils totalisent 102 tronçons verbaux, 156 tronçons sujets, 115 tronçons directs, 298 tronçons indirects, 1487 mots, 3209 syllabes, 6787 phonèmes dont 2302 voyelles brèves, 561 voyelles longues, 445 semi-voyelles /w/ et /y/, 1006 consonnes fricatives, 1092 consonnes plosives, 692 consonnes liquides /r/ et /l/ et 689 consonnes nasales /m/ et /n/. Les pauses ont été étiquetées par un expert<sup>14</sup> en apposant le signe # sur le texte correspondant à la voix naturelle.

Ces phrases ont été lues à une vitesse moyenne (de 10 à 12 phonèmes/ seconde) par un locuteur jordanien, qui n'a reçu aucune consigne particulière afin d'éviter toute influence susceptible d'altérer sa spontanéité. Elles ont été échantillonnées à 16 kHz et analysées à l'aide du logiciel ElanStudio. Après l'alignement des phrases naturelles et synthétiques, ElanStudio permet d'extraire les valeurs de F0 et de la durée des phonèmes en répondant à une requête de l'utilisateur. Celle-ci admet plusieurs critères, dont les plus importants sont : nature du segment analysé (demi-phonème, phonème, syllabe, mot) ; pour les phonèmes, nom du phonème, mode articulaire (fricative, plosive, liquide, glottale et semi-voyelle), trait de voisement (voisée/non voisée), contexte phonétique, nature de la syllabe (ouverte/fermée, accentuée/non accentuée), nombre de phonèmes dans le mot, nombre de syllabes dans le mot,

<sup>14</sup> Mr Zemirli, spécialiste en linguistique computationnelle à l'Institut National d'Informatique d'Alger.

nombre de syllabes dans le groupe de souffle, position du phonème dans le mot, modalité de la phrase, etc. La figure 25 représente l'interface graphique d'ElanStudio.

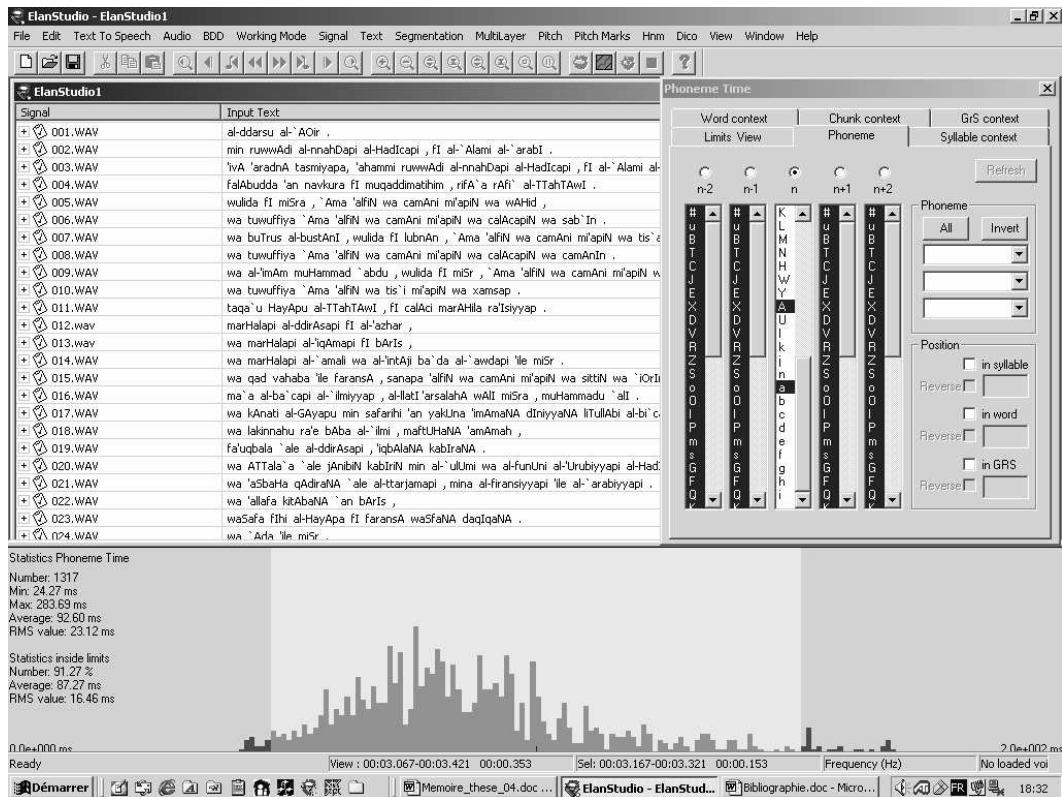


Fig. 25 : Interface graphique d'ElanStudio.

La fenêtre en haut à gauche présente le corpus de phrases, la fenêtre en haut à droite présente les critères de recherche sélectionnés et la fenêtre du bas les résultats de la recherche. Ces derniers sont donnés sous forme d'histogramme et de valeurs statistiques (nombre d'unités extraites, moyenne et écart-type).

## 8.2. Gestion des pauses

Cette section présente le modèle de prédiction automatique de la position des pauses dans le cadre d'un système de synthèse de la parole. Ce modèle est basé sur les signes de ponctuation et sur des considérations phonotactiques, comme le nombre maximal de syllabes non séparées par une pause, en adéquation avec la structure syntaxique de l'énoncé. Ainsi, une pause n'est jamais insérée à l'intérieur d'un tronçon, issu de l'analyse linguistique, mais peut l'être à une frontière syntaxique si le nombre de syllabes depuis la dernière pause est suffisant.

Lorsque nous parlons, nous produisons des sons ponctués de pauses à des intervalles plus ou moins réguliers. Une marque de pause laisse le temps à l'auditeur d'assimiler et d'interpréter le message reçu. À l'inverse, l'absence de pause augmente sa charge de concentration, ce qui entraîne une lassitude puis un désintérêt de celui-ci. Au niveau de la lecture d'un énoncé, nous distinguons les pauses associées aux indicateurs de surface, symbolisés par les signes de ponctuation (point final, virgule, accolade, etc.), des pauses marquées par le lecteur à certaines positions du texte, sans que celles-ci ne soient attestées par des règles connues. Situées généralement aux frontières de groupes syntaxiques, ces pauses sont régies par des contraintes physiologiques pesant sur la phonation et la respiration.

Pour un système de synthèse de la parole à partir du texte, générer des pauses est indispensable à l'intelligibilité de la parole synthétique. Celles-ci contribuent à l'amélioration du confort d'écoute et accroissent ainsi la qualité globale du système. Nous verrons que, d'une part, la prédiction de ces pauses sur la base des seuls signes de ponctuation constitue un traitement minimal, mais insuffisant, et d'autre part, que la seule intégration des autres paramètres syntaxiques et phonotactiques ne suffit pas à une bonne prédiction automatique des pauses. Aussi, la durée d'une pause est fonction de son type : pause de fin de phrase, pause à l'intérieur d'une phrase associée à un signe de ponctuation et pause non associée à un signe de ponctuation.

L'observation du corpus d'analyse nous a permis de formaliser des règles qui prédisent la position des pauses dans la phrase. Nous avons comptabilisé 333 marques de pause dans ce corpus, réparties comme indiqué dans le tableau 20.

	Avec signe de ponctuation	Sans signe de ponctuation
Nombre	229	104
Pourcentage	68,77	31,23

**Tab. 20 : Distribution des pauses.**

La première remarque que nous pouvons faire est que le nombre de pauses liées aux signes de ponctuation est près de deux fois supérieur au nombre de pauses ne correspondant à

aucun indicateur de surface. Mais ce rapport ne peut pas être généralisé à l'ensemble des énoncés arabes car ces pourcentages sont étroitement liés aux caractéristiques structurelles du corpus (nombre et longueur des phrases, etc.) et au style de lecture adopté (débit de parole, etc.).

### 8.2.1 Les pauses associées aux signes de ponctuation

Ces pauses sont facilement prédictibles et leur nombre est quasiment stable d'une lecture à une autre. Le *grand groupe de souffle* désigne l'intervalle séparant deux pauses associées à deux signes de ponctuation successifs. Nous distinguons dans cette catégorie les pauses finales, associées au point final, des pauses à l'intérieur des phrases qui correspondent à la virgule, à la parenthèse fermante, etc.

Généralement, une pause est associée à chaque signe de ponctuation dans un texte. C'est ce qui ressort de l'analyse de notre corpus qui compte 86 points finaux, 134 virgules et 9 deux-points. Les indicateurs de surface rencontrés dans les textes courants sont présentés dans le tableau 21.

.	,	;	:	{ }	[ ]	!	« »	?
---	---	---	---	-----	-----	---	-----	---

Tab. 21 : Indicateurs de surface.

#### Règle 1

Une pause est toujours associée à une frontière #FM ou #FT (signe de ponctuation).

Cependant, l'omission d'une marque de pause devant un signe de ponctuation n'a pas toujours d'incidence sur l'énoncé. Exemple : لِأَنَّ لَهَا مَكَانَتَهَا فِي التَّارِيخِ الْعَرَبِيِّ. /bal li?anna lahA makAnatahA fit tArIxiI carabi/ (« ceci, parce qu'elle a sa place dans l'histoire arabe »).

Dans cet exemple, l'omission de la pause associée à la virgule n'a aucune conséquence sur l'énoncé. L'explication que nous pouvons en faire est que cette virgule se situe à une syllabe seulement du début de la phrase, ce qui constitue un intervalle court pour s'arrêter.

### 8.2.2. Les pauses non associées aux signes de ponctuation

Contrairement à la première catégorie de pauses, celles-ci ne sont pas indiquées dans le texte, ce qui rend leur prédiction plus difficile. Aucun modèle ne peut prédire leur position de manière précise à cause de leur nature **non déterministe**. Les connaissances nécessaires à leur calcul sont multiples et de nature différente (phonotactique, syntaxique, sémantique, etc.). Ne pas disposer de l'ensemble de ces connaissances ne permet d'approcher que partiellement les phénomènes qui président à la distribution de ces pauses, avec deux risques possibles : le placement d'une pause à une position interdite ou l'omission d'une pause à une position potentielle.



Par opposition au grand groupe de souffle, le *petit groupe de souffle* désigne l'intervalle qui sépare une pause non associée à un signe de ponctuation de la pause précédente ou suivante. Généralement, ces pauses interviennent à une frontière de groupe syntaxique. Le tableau 22 représente la distribution des pauses non associées aux signes de ponctuations par rapport aux frontières de tronçons. Nous notons toutefois que 10,47% de ces pauses apparaissent à l'intérieur des tronçons.

	À l'intérieur d'un tronçon	À la frontière d'un tronçon
<b>Nombre</b>	11	94
<b>Pourcentage</b>	10,47	89,53

Tab. 22 : Distribution des pauses non associées aux signes de ponctuation.

Exemples :

- À l'intérieur du tronçon

بَلْ، لَأَنَّ لَهَا مَكَانَتَهَا فِي التَّارِيخِ الْعَرَبِيِّ #الإسلاميَّ أَيْضًا  
/bal li?anna lahA makAnatahA fit tArIxil earabi # ?al?islAmiI ?ayDan/ (« ceci, parce qu'elle a sa place dans l'histoire arabe # musulmane également »).

ثُمَّ #عَيْنَ مُدْرِّسًا لِلأَدَبِ وَ التَّارِيخِ  
/cumma euyyina mudarrisan lil?adabi wat tArIXi/ (« ensuite # il a été désigné enseignant de littérature et d'histoire »)

Dans le premier exemple, une pause est insérée à l'intérieur de la séquence du tronçon indirect (في التَّارِيخِ الْعَرَبِيِّ). Ceci pourrait s'expliquer par le dépassement du seuil critique correspondant au nombre de syllabes sans pause (cette pause est insérée à 18 syllabes du début). Dans le second exemple, la pause est marquée à l'intérieur du tronçon verbal (ثُمَّ عَيْنَ), à seulement deux syllabes du début. L'explication la plus plausible que nous pouvons donner est que cette pause est liée au style de lecture employé.

- À la frontière du tronçon

وَأَصْبَحَ قَادِرًا عَلَى التَّرْجَمَةِ، مِنْ الْفَرَنْسِيَّةِ #إلى الْعَرَبِيَّةِ  
/wa ?aSbaha qAdiran ealat tarjamapi minal firansiyyapi # ?ilal earabiyyapi/ (« et il a acquit la capacité de traduire du français # vers l'arabe »)

فَلأَبْدُ أَنْ نَذْكُرَ #في مُقَدِّمَتِهِمْ  
/falAbudda ?an navkura # fl muqaddimatihim/ (« et il faut citer # parmi les premiers »)

Dans ces deux exemples, les pauses sont respectivement marquées à la frontière des tronçons indirect/indirect après 7 syllabes de la virgule et verbal/indirect après 8 syllabes du début.

Hormis les exceptions recensées dans le corpus, la frontière d'un tronçon semble constituer une position potentielle pour un modèle de prédiction des pauses.

## Règle 2

Une pause n'est jamais insérée à l'intérieur d'un tronçon

Nous nous penchons maintenant sur les contraintes phonotactiques qui commandent la distribution des pauses non associées aux signes de ponctuation. Ces contraintes peuvent être formulées en termes de seuils qui expriment les nombres de syllabes minimum et maximum non séparés par une pause.

La figure 26 représente la distribution de la taille des petits groupes de souffle dans le corpus (en nombre de syllabes). Nous constatons que les groupes d'une taille égale à 4 syllabes sont les plus fréquents (21 occurrences) et que la taille de la majorité des groupes varie entre 3 et 8 syllabes. La taille moyenne des petits groupes de souffle (notée moy) calculée sur tout le corpus est de 7,066 syllabes, avec un écart-type (noté  $\sigma$ ) de 3,15 syllabes.

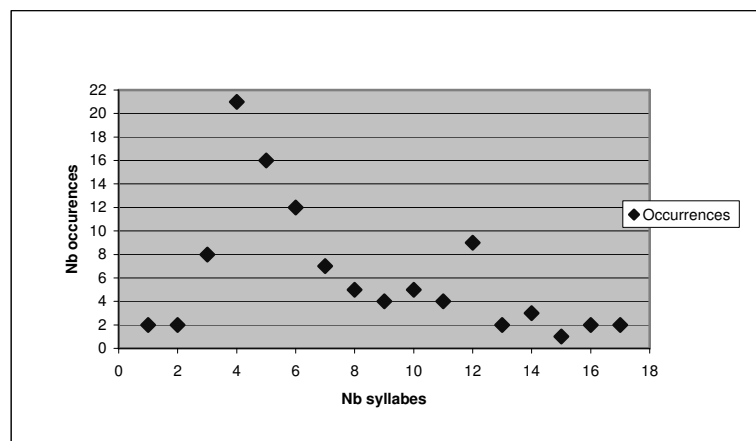


Fig. 26 : Distribution de la taille des petits groupes de souffle.

Nous avons défini à partir de ces résultats et de manière empirique trois seuils qui rendent compte des contraintes de la phonation :

- Le premier seuil1 désigne l'intervalle au-dessous duquel aucune pause ne doit être insérée :  $\text{seuil1} = \text{moy} - \sigma = 3,9 \approx 4$  syllabes.
- Le deuxième seuil2 désigne l'intervalle au-dessus duquel une pause peut être insérée :  $\text{seuil2} = \text{moy} \approx 8$  syllabes.
- Le dernier seuil3 désigne le seuil critique au-dessus duquel une pause doit être insérée :  $\text{seuil3} = 2 * \text{moy} = 16$  syllabes.

Nous avons ensuite étudié la corrélation entre les types de tronçon et les positions des pauses. Autrement dit, une pause peut-elle être insérée à n'importe quelle frontière de tronçon si les contraintes phonotactiques le permettent, ou bien existe-t-il des frontières de tronçon plus susceptibles que d'autres de recevoir une pause ? Pour répondre à cette question, nous avons relevé la distribution des pauses par type de tronçon dans le tableau 23 dont la première colonne (resp. ligne) représente le tronçon précédant la pause (resp. suivant la pause). Seules les pauses à la frontière des tronçons sont représentées.

	Tronçon verbal	Tronçon sujet	Tronçon direct	Tronçon indirect
Tronçon verbal	0	1	3	8
Tronçon sujet	1	2	2	7
Tronçon direct	2	5	1	5
Tronçon indirect	15	0	5	19
Total	18	8	11	39

**Tab. 23 : Distribution des pauses par type de tronçon.**

À la lecture de ce tableau, nous constatons que la majorité des pauses précède un tronçon indirect (39). Par ailleurs, certaines séquences de tronçons ne sont jamais (valeur 0) ou très rarement (valeur 1 ou 2) séparées par une pause. Néanmoins, nous ne pouvons pas tirer de règles strictes concernant ces derniers cas car les résultats obtenus sont étroitement liés au corpus d'analyse et au style de lecture.

En plus de cette analyse statistique de la corrélation entre la position des pauses et le type de tronçon, nous avons introduit une autre considération de nature syntaxique qui concerne les phrases verbales. Celles-ci commencent par un verbe qui, ainsi, détermine le début d'une nouvelle phrase (à exclure les phrases imbriquées commençant par un verbe). Par conséquent, la frontière introduisant un tronçon verbal constitue l'emplacement idéal pour l'insertion d'une pause.

Exemple :

أودُّ أَنْ أَطْلُبَ سَيَّارَةَ أَجْرَةٍ # تَأْتِي غَدًا صَبَاحًا فِي سَاعَةٍ مُبَكَّرَةٍ.  
↑  
pause

/ʔawaddu ʔan ʔaTluba sayyArata ʔujratin # taʔti Gadan SabAHan fl sAcatin mubakkaratin/  
 (« je souhaiterais demander un taxi # pour demain matin de bonne heure »)

À partir de ces contraintes rythmiques et syntaxiques, nous ajoutons de nouvelles règles de gestion des pauses. Rappelons que #FM est associé à un signe de ponctuation faible (virgule, parenthèse ouvrante, etc.), #FT à un signe de ponctuation fort (point final) et #fm à une frontière de tronçon. Une pause est insérée à une frontière #fm si :

### Règle 3

- Le nombre de syllabes depuis la dernière pause est supérieur à seuil2 (8 syllabes).

ET

- Le nombre de syllabes jusqu'à une frontière #FM ou #FT suivante est supérieur à seuil1 (4 syllabes).

ET

- Le tronçon suivant est de type indirect (c'est-à-dire introduit par une préposition).

### Règle 4

- Le nombre de syllabes depuis la dernière pause est supérieur à seuil2 (8 syllabes).

ET

- Le nombre de syllabes jusqu'à une frontière #FM ou #FT suivante est supérieur à seuil1 (4 syllabes).

ET

- La frontière sépare un tronçon objet ou indirect et un tronçon verbal.

### Règle 5

- Le nombre de syllabes depuis la dernière pause est supérieur au seuil critique seuil3 (16 syllabes).

Exemple :

(#FT) (فوق الطاولة) (#P) (#fm) (محفظته) (#fm) (الولد الصَّغِير) (#fm) (وضع)  
Nb syllabes ( 3 + 3 ) 14 = ( 5 ) + ( 3 + 3 ) + ( 3 )

Dans cette phrase, la pause (#P) est insérée avant le tronçon indirect. À cette position, seuil2 = 8 syllabes est dépassé (14 syllabes) et la distance (6 syllabes) séparant (#P) de la marque (#FT) est supérieure à seuil1 = 4 syllabes. La règle 3 est donc appliquée.

#### 8.2.3. Durée des pauses

La durée des pauses est un indice important en synthèse de la parole à partir du texte qui contribue à la compréhension des énoncés. Elle renseigne sur le degré de rupture entre les différents constituants de la phrase : une durée longue signifie une rupture totale, comme la fin d'une section ou d'une phrase, alors qu'une durée courte traduit une rupture partielle, comme la virgule, les deux-points, etc.

Dans son article sur la modélisation de la durée des pauses, Vannier [Van99b] quantifie les forces des frontières syntaxiques des différents constituants de la phrase (distance entre frontières), puis les met en rapport avec les durées des pauses. Le tagging et la mise en relation implémentés sont décrits dans Vergne [Ver98b]. Ainsi, le calcul de ces durées est tributaire des valeurs issues de l'analyse syntaxique. Nous proposons pour notre part de hiérarchiser les durées des pauses selon leur type, puis d'associer à chacune des classes une durée moyenne. Ce choix est guidé par un souci de simplification.

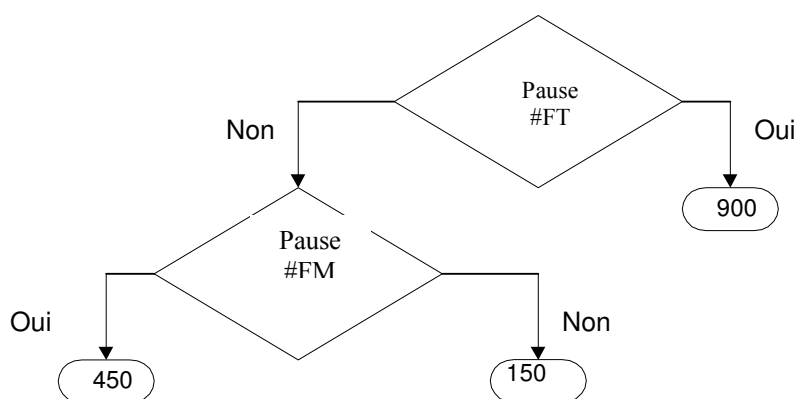


Fig. 27 : Hiérarchisation des durées des pauses.

Comme le montre la figure 27, nous distinguons les pauses terminales associées à #FT des pauses non terminales d'une part, et les pauses intermédiaires #FM associées aux signes de ponctuation des pauses intermédiaires #fm non associées aux signes de ponctuation d'autre part. Les durées de ces pauses ont été obtenues de manière empirique et sont exprimées en millisecondes.

#### 8.2.4. Limites du modèle de gestion des pauses

Le modèle de gestion des pauses que nous proposons est théoriquement en parfaite cohérence avec les unités syntaxiques que représentent les tronçons. Bien que ce modèle ne permette pas de déceler l'emplacement de toutes les pauses dans une phrase, nous avons essayé de nous conformer à la règle selon laquelle l'omission d'une pause affecte moins la cohérence de l'énoncé que l'ajout d'une pause à un endroit interdit. Par exemple, une pause n'est pas systématiquement insérée à une frontière de tronçon quand le seuil minimal de 8 syllabes est atteint, mais seulement si le seuil critique de 16 syllabes est dépassé.

L'usage de ce modèle de pause est directement lié aux résultats issus de l'analyse morpho-syntaxique qui segmente la phrase en tronçons. Une frontière erronée ou une erreur sur les types de tronçon peut **incontestablement** conduire à une mauvaise répartition des pauses, en particulier les erreurs sur le tronçon verbal qui constitue une frontière potentielle pour l'insertion d'une pause. Les frontières du tronçon indirect sont, pour leur part, plus facilement détectables.

Une autre limite de ce modèle est que les connaissances syntaxiques qu'il utilise ne rendent pas compte des relations entre tronçons. En effet, les frontières de constituants syntaxiques de niveau supérieur au tronçon constituent des cibles plus stables pour l'insertion des pauses, notamment dans le cas de phrases imbriquées. Pour illustrer ces propos, examinons l'exemple suivant :

الوَلَدُ الَّذِي حَصَلَ عَلَى الْمِنْحَةِ مَسْرُورٌ  
/?alwaladu ?allavI HaSSala calal minHati masrUrun/ (« l'enfant qui a obtenu le prix est content »)

Cette phrase nominale renferme la phrase imbriquée *الَّذِي حَصَلَ عَلَى الْمِنْحَةِ* /?allavI HaSSala calal minHati/ (« qui a obtenu le prix ») qui a la valeur d'épithète. Une application des règles de notre modèle placerait une pause avant le tronçon indirect *عَلَى الْمِنْحَةِ* /calal minhati/ (à 9 syllabes du début de la phrase), à l'intérieur donc de la phrase de second niveau, ce qui peut altérer le sens général de l'énoncé.

Une mise en relation des tronçons permettrait de privilégier les frontières du niveau le plus élevé pour le placement des pauses. Ainsi, comme le montre la figure 28, la pause serait insérée à la fin de la phrase imbriquée (les crochets représentent le niveau le plus élevé) à l'issue de la liaison des constituants.

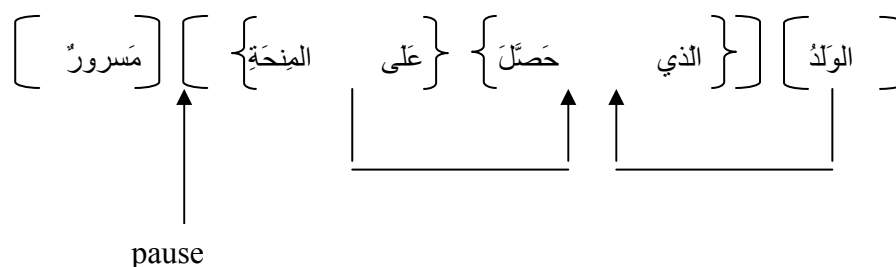


Fig. 28 : Exemple d'insertion d'une pause après la mise en relation des tronçons.

Pour pallier ce problème, nous avons développé des solutions ad hoc en réinitialisant par exemple le compteur de syllabes parcourues dès la rencontre d'un pronom relatif. Mais la gestion des pauses reste un phénomène complexe qui obéit à des contraintes diverses de nature rythmique, syntaxique et parfois même sémantique. De plus, un même texte lu par un même ou plusieurs locuteurs peut être scandé de manière différente. Le tronçon constitue un premier niveau de connaissance qui contribue à une modélisation acceptable de la distribution des pauses, mais il faudra sans doute accéder à des niveaux supérieurs pour une meilleure approche des phénomènes à l'origine des pauses.

### 8.3. Génération de la fréquence fondamentale

La modélisation de la courbe de la fréquence fondamentale est une opération fort complexe de part les phénomènes qui la gouvernent. L'origine de ces phénomènes est diverse : morphologique, syntaxique, phonotactique, etc. Le premier rôle d'un modèle de génération de F0 est de rendre compte des faits intonatifs généraux comme la déclinaison, la réinitialisation de la courbe mélodique ou son abaissement (resp. élévation) à la fin des phrases déclaratives (resp. interrogatives). Les autres faits, comme la corrélation à la syntaxe, sont beaucoup moins évidents à reproduire car difficilement déchiffrables dans les corpus.

Sous l'angle de la syntaxe, l'analyse d'un corpus de phrases à des fins de modélisation prosodique pose deux types de problème : d'abord, un corpus est un ensemble fini de structures syntaxiques, ce qui limite le champ d'analyse aux seuls faits présents. Ensuite, et compte tenu de la diversité de ces structures, il est difficile de dégager des règles de comportement générique qui s'appliquent à l'ensemble des phrases. Dans ce cas, l'analyse doit se focaliser sur les événements les plus répandus.

Nous proposons dans ce travail deux approches pour la modélisation de la courbe mélodique. La première approche est fondée sur le mot : nous avons tracé la droite qui s'ajuste au mieux avec les valeurs de F0 portée par les syllabes accentuées des mots ; la seconde approche est fondée sur les tronçons que nous avons définis au chapitre précédent comme des syntagmes non récursifs : nous appelons *accent de tronçon* l'accent réalisé comme le pic mélodique du tronçon. Nous avons alors tracé la meilleure droite qui s'ajuste au mieux avec ces accents de tronçon.

#### 8.3.1. Analyse de l'accent lexical

##### A. Règles d'assignation de l'accent lexical

Dans ce qui suit, nous désignons par accent lexical ce que d'autres auteurs décrivent comme l'accent primaire. Il a été suggéré que la détection cet accent est un point d'ancrage suffisant pour la génération de l'intonation arabe [Han00] [Zak00a]. Les règles d'assignation de l'accent lexical présentées dans la section 6.2.2 divergent sur deux points : la remontée de l'accent lexical au-delà de l'antépénultième et l'accentuation des sur-lourdes en fin de mot. En considérant que le maximum intonatif dans le mot se réalise sur la syllabe accentuée, nous avons analysé un corpus de 60 mots isolés afin de vérifier la validité de ces règles du point de vue acoustique. Deux valeurs de F0 sont prélevées sur chaque voyelle (un F0 moyen par demi-voyelle- les F0 sont calculés par intervalle de 5 ms). Ces valeurs sont exprimées en demi-ton (DT), par rapport à la référence 50 Hz. Par exemple, la valeur de 12 DT est équivalente à 100 Hz (12 DT = 6 Tons = 1 octave =  $2 \times 50$  Hz).

Nous présentons ci-dessous quelques exemples de contours mélodiques qui illustrent les tendances de F0 au niveau du mot.

1. Mots dont la structure syllabique est /CV1CV2CV3/

	CV1		CV2		CV3	
/faqada/	20,39	20,52	17,47	15,17	12,19	10,50
/daraka/	17,86	19,18	18,07	16,41	12,48	10,36
/vahaba/	16,18	19,07	17,21	14,70	12,38	8,88

Dans ces exemples, nous notons que le maximum mélodique se réalise sur la syllabe définie comme accentuée par El-Ani. Il est atteint dans la deuxième partie de la voyelle. Ainsi, le contour intonatif monte de la première syllabe FD jusqu'à la syllabe accentuée FA, puis diminue de la syllabe FA jusqu'à la syllabe finale FF. Le minimum mélodique se réalise sur la dernière syllabe FF.

2. Mots dont la structure syllabique est /CV1CV2C3V3C4V4/ où CV1 correspond au préfixe /fa/

	CV1		CV2		CV3		CV4	
/favahaba/	13,21	13,07	18,38	18,49	14,23	12,97	13,81	13,05
/fafaqada/	14,04	13,16	18,26	20,45	17,63	15,04	11,38	10,85

Nous notons que le maximum mélodique se réalise sur la syllabe accentuée CV2 définie comme accentuée par El-Ani, ce qui indique que l'introduction d'un préfixe n'a pas affecté la tendance du contour intonatif. Celui-ci monte de FD à FA, puis diminue de FA à FF.

3. Mots dont la structure syllabique est /CVC1CV2CV3CV4/

Nous cherchons ici à étudier la remontée de l'accent lexical au-delà de l'antépénultième CV2. En d'autres termes, vérifier si la syllabe CVC1 est susceptible de recevoir l'accent lexical.

	CVC1		CV2		CV3		CV4	
/mafcalata/	19,52	20,69	19,8	19,09	14,97	15,82	15,23	12,58
/maskanahu/	17,39	20,72	18,52	17,114	15,01	14,38	16,17	13,95
/maslakahu/	18,67	26,01	19,1	19,47	15,2	14,52	16,01	13,85
/baeva?vin/	17,35	19,70	15,72	16,52	15,86	14,48	16,10	12,54

Nous notons que la valeur de F0 est maximale sur la première syllabe CVC1 (deuxième partie de la voyelle) et minimale sur la dernière syllabe. Ce phénomène a été observé pour l'ensemble des 15 mots de cette structure. Ceci suggère que l'accent lexical remonte au-delà de l'antépénultième du point de vue acoustique.

4. Mots dont la structure syllabique est CV1CVVC2



Nous cherchons ici à vérifier si la syllabe finale sur-lourde CVVC2 est susceptible de recevoir l'accent lexical.

	CV1		CVVC2	
/naOIT/	13,11	11,48	15,89	11,64
/zamAn/	13,21	12,55	15,66	11,84
/manAr/	11,84	12,13	14,65	10,51
/SaGIr/	13,44	12,22	15,90	10,06

Nous notons que le maximum mélodique se réalise dans la première partie de la voyelle longue CVVC2, c'est à dire sur la syllabe sur-lourde.

### 5. Mots dont la structure syllabique est CV1CVVC2

Ces mots se terminent par la syllabe sur-lourde CVVC

	CV		CVCC	
/Darabn./	11,90	12,56	16,40	15,87
/darasn/	12,70	13,45	16,78	16,40
/laeibn/	11,64	11,64	15,30	16
/Sadamn./	15,55	12,72	15,88	14,68

La même observation est faite pour ces mots : le maximum mélodique se réalise sur la syllabe sur-lourde CVCC. Ceci suggère que l'accent lexical se réalise sur les sur-lourdes dans une position finale du point de vue acoustique. Ce qui va à l'encontre des résultats de El-Ani [Ela70] et Rajouani [Raj89] et conforte la position de Kouloughli [Kou76].

Nous proposons les règles suivantes pour la détermination de la position de l'accent lexical :

- a. Si la dernière syllabe du mot est une sur-lourde, alors elle porte l'accent lexical.
- b. Si la règle (a) ne s'applique pas et si le mot comporte une ou plusieurs syllabes longues (lourde ou sur-lourde) alors la syllabe longue la plus proche de la fin du mot porte l'accent lexical. Dans ce cas, la dernière syllabe est ignorée (extra-métrique).
- c. Si le mot est constitué uniquement de syllabes de type CV, la première syllabe porte alors l'accent primaire. Le préfixe est dans ce cas ignoré.

La validation de ces règles se base sur l'analyse des variations de la fréquence fondamentale au niveau du mot. Étant donné que notre but est de reproduire ces variations, la position de l'accent lexical est définie du point de vue des paramètres physiques. Sous l'angle

de la linguistique, et compte tenu des divergences au sujet de l'accent, il ne nous appartient pas à notre niveau de prendre position.

## B. Validité des règles d'assignation de l'accent lexical au niveau de la phrase

Pour vérifier si la place de l'accent lexical est sauvegardée dans la phrase, nous avons étudié l'évolution du contour mélodique au niveau de la phrase en prélevant les valeurs de F0 sur l'ensemble des voyelles (un F0 moyen par voyelle). La notation suivante est utilisée :

- Le signe → sous la syllabe de rang n indique que la variation de F0 entre la syllabe n-1 et n ne dépasse pas 1 DT. Il signifie une stabilité (ST) de F0.
- Le signe ↑ indique la montée (MT) de F0 d'une valeur supérieure à 1 DT.
- Le signe ↓ indique la descente (DT) de F0 d'une valeur supérieure à 1 DT.

الدرس الخامس /?addarsul xAmis/

	?ad	<b>dar</b>	sul	<b>xA</b>	Mis
F0	9,65	14,05	13,31	10,22	4,95
MT/DT/ST	↑	↑	→	↓	↓

ويبلغ عدد سكانها اليوم /wa yabluGu eadadu sukkAnihal yawm/

	wa	<b>yab</b>	lu	Gu	<b>ea</b>	da	du	suk	<b>kA</b>	ni	hal	<b>yawm</b>
F0	9,22	17,76	16,39	14,75	13,32	12,01	9,26	10,61	12,89	6,86	6,61	10,68
MT/DT/ST	↑	↑	↓	↓	↑	↓	↓	↓	↑	↓	↓	↑

فقد كانت مركز الحضارة الإسلامية /faqad kAnat markazal HaDAratil ?islAmiyyap/

	Fa	qad	<b>kA</b>	nat	<b>mar</b>	ka	zal	Ha	<b>DA</b>	ra	pil	?is	lA	<b>miy</b>	ya
F0	11,13	14,20	19,63	18,82	17,58	16,68	13,88	12,11	15,03	14,57	13,75	11,84	10,81	13,10	9,81
MT/DT/ST	↑	↑	↑	→	↓	→	↓	↓	↑	→	→	↓	↓	↑	↓

Au niveau des mots, nous observons que le maximum mélodique se réalise toujours sur la syllabe accentuée (en gras). Par conséquent, la position de l'accent lexical est préservée dans la phrase.

### 8.3.2. Analyse de l'intonation

Cette étude repose sur l'extraction des valeurs moyennes de F0 sur les voyelles des syllabes accentuées de la voix naturelle : d'abord, sans distinction de longueur, de structure syntaxique des phrases ; ensuite selon la longueur des phrases, leur modalité et leur

organisation en tronçons. Dans ce dernier cas, le but recherché est de dégager des tendances de la courbe mélodique en fonction des critères retenus.

Etant donné la diversité structurelle des phrases dans le corpus, nous ne présenterons pas les résultats selon des critères syntaxiques. Sans les reproduire dans leur ensemble, nous considérons dans ce qui suit quelques exemples de phrases qui illustrent les tendances de la courbe mélodique. Nous dégagerons des règles préliminaires sur les variations de F0, puis nous étudierons leur validité sur l'ensemble du corpus.

#### A. Cas de la phrase déclarative

Nous présentons pour chaque phrase, en plus des valeurs de F0, les frontières et les types de tronçons : TV (tronçon verbal), TS (tronçon sujet), TD (tronçon direct), TI (tronçon indirect). Les parenthèses ouvrante et fermante désignent respectivement les débuts et fins de tronçon et le symbole # la position de la pause dans la phrase (P).

##### 1. lakinnahu ra?e bAba al-eilmi , maftUHaN ?amAmah.

Tronçon	TV	TV	TD		P	TD	TI
Phrase	lakinnahu	Ra?e	bAba	al-eilmi	#	maftUHaN	?amAmah
F0	17,50	16,52	14,69	13,72		15,31	9,47
MT/DT/ST	↑	→	↓	→		↑	↓

Cette phrase est constituée de deux groupes de souffle (GRS) séparés par une pause. Dans les deux GRS, nous constatons une montée intonative sur le premier mot, puis une descente intonative jusqu'au dernier mot. À l'exception du tronçon TD qui précède la pause dans le premier GRS, les autres tronçons se composent d'un seul mot. La pause ici est accompagnée d'une réinitialisation de F0.

##### 2. taqaœu HayApu al-TTahTAwI fI calAci marAHila ra'Isiyyap.

Tronçon	TV	TS		P	TI			
Phrase	taqaœu	HayApu	al-TTahTAwI	#	fI	calAci	marAHila	ra'Isiyyap
F0	20,89	17,79	8,75		13	18,71	15,8	9,15
MT/DT/ST	↑	↓	↓		↑	↑	↓	↓

Cette phrase est constituée de deux GRS. Dans le premier GRS, nous constatons une montée intonative sur le premier mot, puis une descente sur les deux derniers mots du GRS. En d'autres termes, une montée de F0 dans le tronçon verbal, puis une descente dans le tronçon sujet, qui est le dernier tronçon du GRS. Dans le deuxième GRS, nous notons une montée de F0 jusqu'au deuxième mot du tronçon indirect, puis une descente sur les deux

derniers mots du tronçon : donc, une montée de F0 jusqu'au premier mot lexical du tronçon indirect (fl est un mot grammatical), puis une descente dans ce même tronçon sur les deux derniers mots du GRS (ce tronçon est le dernier du GRS).

3. wulida fl miSra εAma ?alfiN wa camAni mi?apiN wa wAHid.

Tronçon	TV	TI		TI		TC			TC	
Phrase	wulida	fl	miSra	εAma	?alfiN	wa	camAni	mi?apiN	wa	wAHid
F0	15,87	0,38	13,39	8,04	11,96	0,36	8,95	11,2	5,45	9,76
MT/DT/ST	↑	↓	↑	↓	↑	↓	↑	↑	↓	↑

Cette phrase est constituée d'un seul GRS composé de 5 tronçons. Nous notons deux tendances du contour mélodique : à l'intérieur des tronçons, une montée de F0 du début à la fin du tronçon ; au niveau de la phrase, une baisse de F0 sur les derniers mots des tronçons du début à la fin de la phrase. Cette tendance de F0 à monter du début à la fin du tronçon n'a pas été constatée dans les phrases 1 et 2. Ceci est peut-être lié, comme il sera examiné plus loin, à la position du tronçon dans la phrase : dans un tronçon final, le F0 baisse du premier mot lexical au dernier mot du GRS.

4. al-llatI tarakat ?acraha fl al-?AdAbi wa al-mUsIqe al-εAlamiyyap.

Tronçon	TV		TD	TI		TC		
Phrase	wa	tarakat	?acraha	fl	al-?AdAbi	wa	al-mUsIqe	al-εAlamiyyap
F0	9,42	14,36	11,59	7,66	11,25	9,09	10,42	5,29
MT/DT/ST	↑	↑	↓	↓	↑	↓	↑	↓

Nous notons une montée de F0 à l'intérieur des tronçons non finaux (du début à la fin du tronçon). Sur le dernier mot des tronçons, le F0 du dernier mot du tronçon baisse du début à la fin de la phrase. Pour le tronçon final TC, le F0 augmente du début au premier mot lexical, puis diminue jusqu'à la fin de la phrase.

5. al-qAhirap , εAma ?alfiN wa tisei mi?apiN wa sabeapiN wa xamsIn .

Tronçon	TS	P	TI		TC			TC		TC	
Phrase	al-qAhirap	#	εAma	?alfiN	wa	tisei	mi?apiN	wa	sabeapiN	wa	xamsIn
F0	12,43		9,35	13,96	8,98	9,15	13,11	2,14	15,33	6,48	6,88
MT/DT/ST	↑		↑	↑	↓	→	↑	↓	↑	↓	→

La tendance précédente, qui se confirme dans la phrase 5, a été observée pour plusieurs phrases dans le corpus. Nous nous sommes intéressés dans les exemples suivants aux tronçons en début de phrase.

6. wa qad Aittabaæa wasA?ila mutaæaddidap .

Tronçon Phrase	TV			TD	
	wa	qad	Aittaba	wasA?ila	mutaæaddidap
F0	9,45	14,42	21,62	17,93	7,62
MT/DT/ST	↑	↑	↑	↓	↓

7. al-funUnu al-?adabiyapu wa ?ælamuhA fl al-nnahDapi al-æarabiyapi al-HadIcap.

Tronçon Phrase	TS		TC		P #	TD			
	al- funUnu	al- ?adabiyapu	wa	?ælamuhA		fl	al- nnahDapi	al- æarabiyapi	al- HadIcap
F0	18,47	15,68	5,73	8,11		9,94	13,96	10,93	10,49
MT/DT/ST	↑	↓	↓	↑		↑	↑	↓	→

8. al-muTAlaæapu al-wAfiyapu lilmadArisi al-ccAnawiyap.

Tronçon Phrase	TS		TI	
	al-muTAlaæap	al-wAfiyap	lilmadArisi	al-ccAnawiyap
F0	14,24	9,12	12,23	12,05
MT/DT/ST	↑	↓	↑	→

L'observation des exemples 4 et 6 indique la montée de F0 du début à la fin du tronçon initial, qui se compose d'un ou de plusieurs mots grammaticaux suivis d'un mot lexical. En conséquence, le F0 est maximal sur le mot lexical du tronçon. Dans les exemples 7 et 8, nous notons une montée de F0 sur le premier mot, puis sa diminution jusqu'à la fin du tronçon. À la différence des précédents, ces tronçons initiaux commencent par un mot lexical sur lequel le F0 est maximal.

Les faits essentiels qui se dégagent de l'analyse du corpus sont récapitulés avec les règles suivantes :

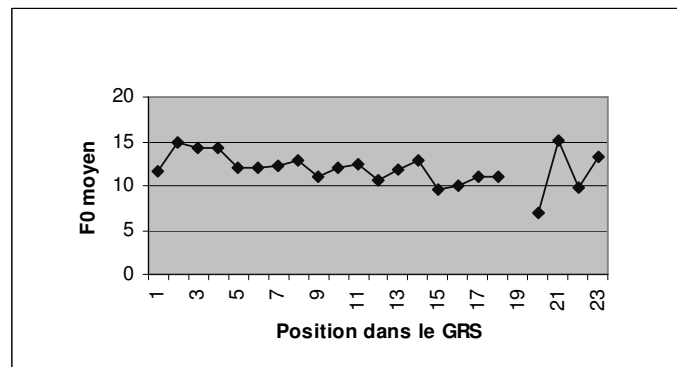
- Dans un tronçon initial, le contour mélodique augmente du début jusqu'au premier mot lexical, puis diminue jusqu'à la fin du tronçon. Le maximum mélodique se réalise sur le premier mot lexical du tronçon.
- Dans un tronçon intermédiaire, le contour mélodique augmente du début à la fin du tronçon. Le maximum mélodique se réalise sur le dernier mot du tronçon.
- Dans un tronçon final, le contour mélodique augmente du début jusqu'au premier mot lexical, puis diminue jusqu'à la fin du tronçon. Le maximum mélodique se réalise sur le premier mot lexical du tronçon.

De part la nature instable de la substance phonétique et la complexité des phénomènes qui la régissent, il existe des phrases dans le corpus qui ne suivent pas les tendances observées. Pour cette raison, il est important de valider les règles retenues sur l'ensemble du

corpus, d'abord en calculant les pentes des phrases en considérant toutes les syllabes accentuées, ensuite en calculant ces mêmes pentes en ne gardant que les valeurs maximales des tronçons. Les coefficients de corrélation des différentes pentes seront ensuite comparés.

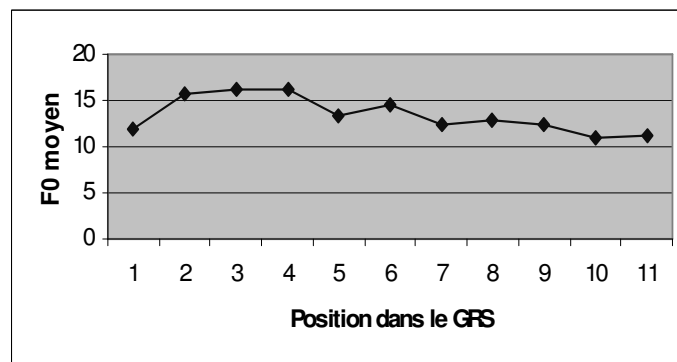
#### A.1. Une approche basée sur le mot

Dans cette approche, les valeurs moyennes de F0 sont prélevées sur les voyelles des syllabes accentuées, en tenant compte de l'ensemble des mots du groupe de souffle GRS. Le choix de l'unité GRS permet de neutraliser les fluctuations de F0 occasionnées par la réinitialisation de la courbe intonative à l'intérieur des phrases. La figure 29 représente l'évolution de F0 en fonction de la position des syllabes accentuées dans le GRS.



**Fig. 29 : Evolution de F0 en fonction de la position des syllabes accentuées dans le GRS.**

En considérant le corpus dans son ensemble, il semble exister une déclinaison de la courbe intonative. La figure 30 représente l'évolution de F0 en fonction de la position des syllabes accentuées dans le GRS de 12 syllabes (GRS majoritaire dans le corpus-116 GRS).

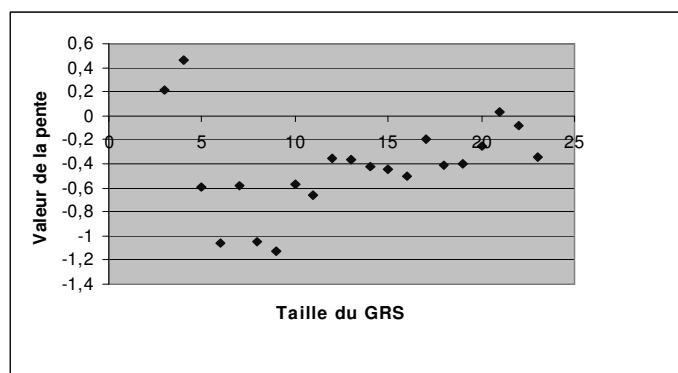


**Fig. 30 : Evolution de F0 en fonction de la position des syllabes accentuées dans le GRS de 12 syllabes.**

Cette figure confirme la tendance de F0 à décroître dans le GRS de 12 syllabes, avec une pente négative égale à  $-0,35$  DT/syllabe. Cette analyse va dans le sens des résultats obtenus pour les autres langues [Bou01] sur l'existence de la déclinaison en arabe. À partir de

là, nous avons calculé les pentes de déclinaison des GRS en fonction de leur taille (cf. figure 31) : une valeur moyenne de pente est associée à chaque taille de GRS mesurée en nombre de syllabes.

Nous avons utilisé la méthode des moindres carrés pour estimer la ligne de déclinaison la plus proche des valeurs de F0. Il s'agit de la ligne haute de déclinaison, c'est-à-dire, la ligne qui passe par les maxima de la courbe intonative qui sont instanciés par les syllabes accentuées.

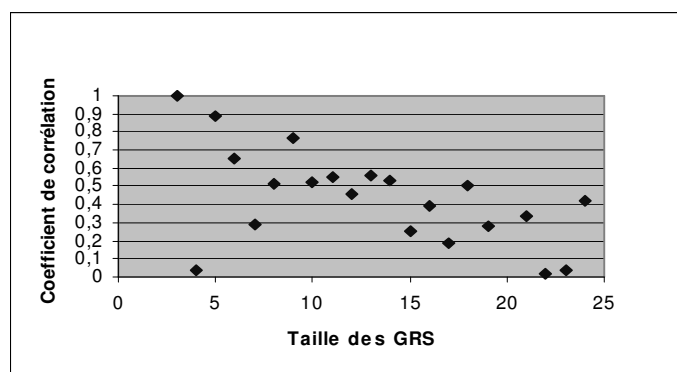


**Fig. 31 : Pente moyenne en fonction de la taille du GRS.**

Une grande partie des valeurs de pente obtenues se situe entre 0 et -0,6. Les pentes les plus importantes sont relevées pour les GRS courts de moins de 10 syllabes, alors qu'elles se rapprochent de zéro pour les GRS de plus de 15 syllabes. Il est cependant difficile de déterminer une tendance de la pente en fonction de la taille des GRS en raison de la disparité des valeurs obtenues. Il existe par ailleurs des valeurs de pentes positives pour les GRS de moins de 4 syllabes qui pourraient correspondre à une montée de F0 à l'intérieur des tronçons.

Mais ces pentes de déclinaison sont-elles valides pour prévoir les valeurs de F0 en fonction de la taille du GRS ? Les droites de régression n'assurent pas toujours une parfaite corrélation avec les valeurs réelles. En conséquence, cette abstraction peut passer outre certaines valeurs de F0.

Pour mesurer la pertinence de ces droites de régression, nous avons calculé les coefficients de corrélation qui comparent les valeurs estimées de F0 aux valeurs réelles. Rappelons qu'un coefficient de corrélation égal à 0 indique que la pente ne peut servir à prévoir les valeurs de F0 et qu'un coefficient égal à 1 indique une parfaite corrélation de ces valeurs. La figure 32 représente les valeurs des coefficients de corrélation en fonction de la taille du GRS.



**Fig. 32 : Coefficient de corrélation en fonction de la taille du GRS.**

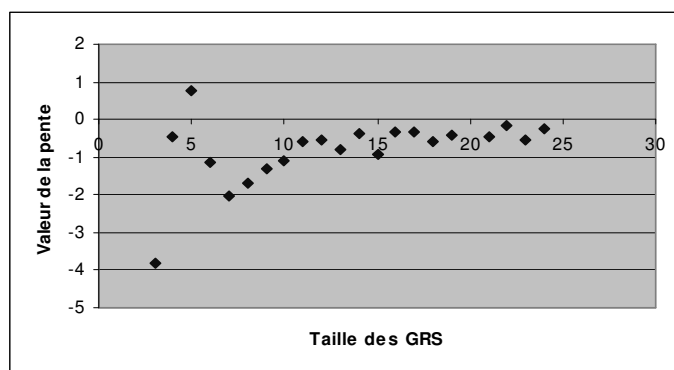
À la lecture de ce graphique, nous relevons que 50% des coefficients de corrélation sont inférieurs à 0,5 d'une part et que 27,27% d'entre eux se situent entre 0,5 et 0,6 d'autre part. Seuls 18,18% sont supérieurs à 0,6. Par conséquent, les lignes de déclinaison ne sont pas satisfaisantes pour prédire les valeurs de F0.

Au vu de ces résultats, une autre question sur la pertinence des valeurs de F0 considérées pour le calcul de la droite de régression peut être posée : avons-nous besoin de considérer les valeurs de l'ensemble des syllabes accentuées pour l'estimation de l'équation de régression ? Dans cette première approche, nous n'avons pris en compte que les phénomènes *accentuels* pour la prédiction des valeurs de F0 en faisant abstraction des autres phénomènes, comme la structure syntaxique des phrases. Ceci peut expliquer les mauvais résultats obtenus.

## A.2. Une approche basée sur le tronçon

Dans cette seconde approche, une seule valeur par tronçon est retenue pour le calcul des droites de régression. Elle correspond à la valeur moyenne de F0 prélevée sur les voyelles des syllabes accentuées selon la position du tronçon : dans un tronçon initial ou final, elle est prélevée sur le premier mot lexical du tronçon ; dans un tronçon intermédiaire, elle est prélevée sur le dernier mot du tronçon. Ces valeurs correspondent aux maxima mélodiques des tronçons comme nous l'avons vu précédemment. Les pentes des droites de régression qui passent par les valeurs de F0 ainsi prélevées sont représentées dans la figure 33 en fonction de la taille du GRS.



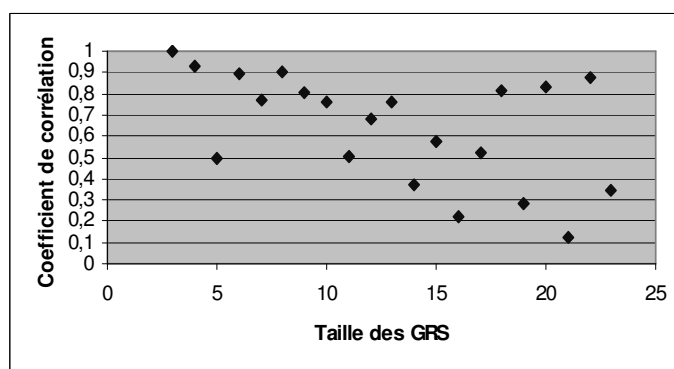


**Fig. 33 : Pente moyenne en fonction de la taille du GRS.**

Nous constatons que ces valeurs de pente sont négatives à l'exception de la valeur calculée pour les GRS de 5 syllabes. Dans leur ensemble, elles sont moins importantes que celles de la première approche. Néanmoins, à l'instar des précédentes, les pentes les plus importantes sont relevées pour les GRS courts de moins de 10 syllabes, et les moins importantes pour les GRS de plus de 10 syllabes. Au delà de 15 syllabes, ces valeurs sont plus stables et se rapprochent de zéro.

Contrairement à l'approche précédente, il se dessine ici une corrélation plus nette entre les valeurs de pente et la taille du GRS (la forme de la courbe est hyperbolique) : plus la taille du GRS est petite, plus la pente est importante. Les valeurs de pente les plus faibles sont relevées pour les phrases longues de plus de 15 syllabes.

Nous avons calculé les coefficients de corrélation pour ces nouvelles valeurs de pente et obtenu les résultats représentés dans la figure 34. Nous notons que 77% de ces coefficients sont supérieurs à 0,5 dont 58% sont au-dessus de 0,7. Ainsi, 23% d'entre eux sont inférieurs à 0,5 contre 50% dans la première approche.



**Fig. 34 : Coefficient de corrélation en fonction de la taille du GRS.**

En observant cette figure, nous constatons que les valeurs de coefficient les plus faibles concernent les GRS de plus de 15 syllabes. À l'inverse, elles sont proches de 1 pour les GRS de moins de 10 syllabes. Ceci peut s'expliquer par le fait que les GRS de taille importante s'accompagnent le plus souvent d'une réinitialisation de la courbe mélodique.

En dernier lieu, nous avons comparé les coefficients de corrélation des deux approches. Nous constatons dans la figure 35 que les coefficients de corrélation de la seconde approche sont majoritairement supérieurs aux coefficients de la première approche. Par conséquent, les valeurs de F0 liées aux tronçons semblent plus *pertinentes* que les valeurs de F0 liées aux mots pour le calcul des droites de régression dans les GRS des phrases déclaratives.

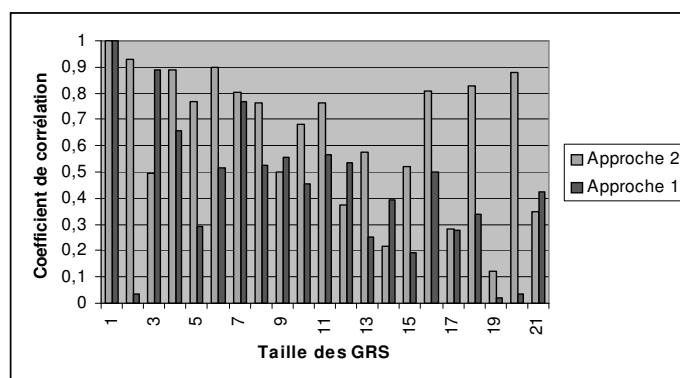


Fig. 35 : Comparaison des coefficients de corrélation des approches 1 et 2 en fonction de la taille du GRS.

En nous appuyant sur la figure 33 qui représente les valeurs moyennes de pente en fonction de la taille du GRS, nous avons retenu la fonction suivante pour le calcul de la ligne haute de déclinaison :

$$Pente = \frac{12}{Nombre\_Syllabe + 2}$$

La ligne de base de déclinaison représente la droite qui passe par les minima de la courbe mélodique. Nous n'avons observé aucune corrélation entre ces minima mélodiques, qui se réalisent le plus souvent sur les dernières syllabes des mots et la structure syntaxique des phrases.

#### B. Cas de la phrase interrogative

Nous distinguons en arabe les questions totales, qui appellent une réponse par *oui/non*, et les questions partielles, les autres types de questions. Les questions totales peuvent être introduites par les particules (أ /?a/, هل/hal/) ou ne contenir aucun marqueur interrogatif. Les questions partielles peuvent être introduites soit par des adverbes qui ont un sens interrogatif

( /kayfa/, /?ayna/, etc.), soit par un pronom interrogatif ( /man/, /mA/, etc.) [Ben93].

Nous présentons ci-dessous les résultats d'analyse de quelques phrases interrogatives qui illustrent la tendance de la courbe intonative pour ces phrases. En plus des valeurs prélevées sur les voyelles des syllabes accentuées, nous avons extrait la valeur de F0 sur la dernière voyelle de la phrase.

1. Phrase sans marqueur interrogatif : /karramal mudIrut tilmlval mujtahida/ ?

Tronçon	TV	TS	TD	
Phrase	karramal	mudIrut	tilmlval	mujtahida
F0	16,73	17,36	13,40	15,51 22,31
MT/DT/ST	↑	→	↓	↓ ↑

Nous constatons que la courbe mélodique augmente à l'intérieur des deux premiers tronçons, diminue à l'intérieur du troisième tronçon puis augmente à l'intérieur du dernier tronçon.

2. Phrase avec la particule /hal/ : /halil qiTArul qAdimu min tUnis muta?axxiruN/

Tronçon	TS		TS	TI		TS	
Phrase	halil	qiTArul	qAdimu	min	tUnis	muta?	axxiruN
F0	17,40	17,37	18,02	15,19	18,86	19,60	22,08
MT/DT/ST	↑	→	→	↓	↑	↑	↑

3. Phrase avec la particule /?a/ : /?akAnal wAlidayni fl Hayrapin/ ?

Nous représentons dans cet exemple la valeur de F0 prélevée sur la voyelle de la particule /?a/.

Tronçon	TV		TS	TI		
Phrase	?akAnal	wAlidayni	fl	Hayrapin		
F0	25,89	24,66	16,52	19,87	19,25	24,69
MT/DT/ST	↑	→	↓	↑	→	↑

Nous constatons dans les exemples 2 et 3 la même tendance que celle observée dans la phrase 1 : la courbe intonative monte à l'intérieur des premiers tronçons, baisse en milieu de phrase puis augmente à l'intérieur du dernier tronçon. Dans ces deux exemples, la valeur de F0 sur les voyelles des particules interrogatives est élevée. Ceci a été observé par Zaki [Zak00b].

4. Phrases avec le marqueur /?ayna / : /?ayna vahaba havA al-masA?/ ?

Tronçon	TV		TS	
Phrase	?ayna	vahaba	havA	al-masA?
F0	18,44	16,28	15,53	13,92 X
MT/DT/ST	↑	↓	→	↓

5. Phrases avec le marqueur /man/ : man qara?a jarIdata al-yawm?

Tronçon Phrase	TV		TD	
	man	qara'a	jarIdapa	al-yawm
F0	18,06	16,76	15,88	13,60
MT/DT/ST	↑	↓	→	↓

La principale différence entre les exemples 4 et 5 et les précédents réside au niveau de l'intonation finale. Pour ces marqueurs interrogatifs, l'intonation est descendante à l'intérieur du dernier tronçon. Les tendances observées dans ces exemples ont été vérifiées sur plusieurs phrases du corpus. Nous les récapitulons dans les règles suivantes :

- Dans les questions totales, le contour mélodique augmente à l'intérieur des premiers tronçons, diminue en milieu de phrase puis augmente à l'intérieur du dernier tronçon.
- Dans les questions partielles, le contour mélodique augmente à l'intérieur du premier tronçon, puis diminue jusqu'à la fin de la phrase.
- Le maximum mélodique du premier tronçon se réalise sur les marqueurs interrogatifs quand ils existent.
- Le maximum mélodique du dernier tronçon se produit sur la dernière syllabe si la question est totale, sinon, sur la dernière syllabe accentuée.

Dans le cas des phrases interrogatives, le contour mélodique global semble moins sensible à la structure syntaxique comparé aux phrases déclaratives. Ce contour dépend du type de marqueur interrogatif employé quand il existe. Ainsi, nous n'avons pas observé de corrélation entre le F0 et la position des tronçons dans la phrase, à l'exception du tronçon final dans lequel le contour de F0 est montant pour les questions totales.

Pour l'estimation du contour mélodique des questions totales, nous avons retenu trois valeurs par phrase : le maximum mélodique du premier tronçon (FM1), le maximum mélodique de l'avant dernier tronçon (FM2) et le maximum mélodique du dernier tronçon (FM3). Les pentes calculées de la ligne de déclinaison qui relie FM1 à FM2 en fonction de la taille des phrases sont représentées dans le tableau 24. À l'intérieur du dernier tronçon, le contour mélodique augmente jusqu'à atteindre FM3 (estimé à 20 DT).

	FM1	FM2	Pente
Taille < 13	18,92	16,08	-0,23
Taille > 12	18,87	16,97	-0,11

Tab. 24 : Pentes de déclinaison en fonction de la taille des phrases.

### 8.3.3. Synthèse de l'intonation

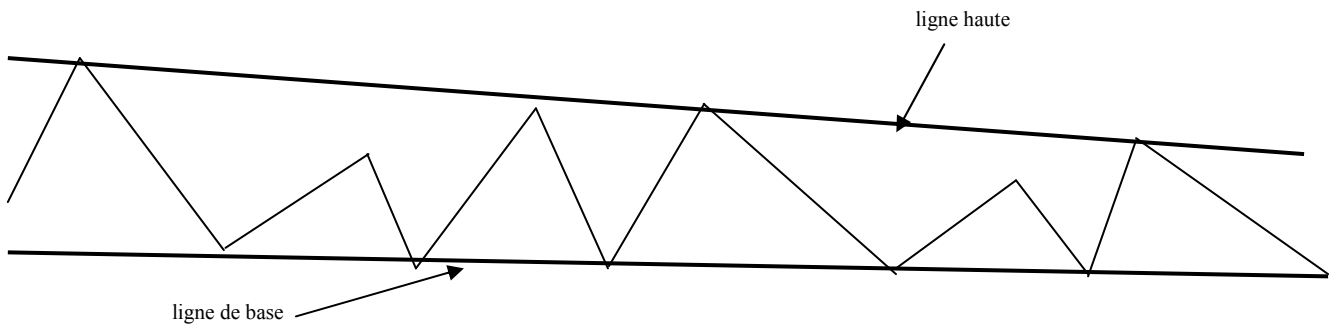
Sur la courbe de F0 d'un mot isolé, le maximum de fréquence se situe sur la syllabe qui porte l'accent lexical. Chaque mot lexical conserve son accent propre dans la phrase. Le contour mélodique dépend de la modalité de la phrase (déclarative ou interrogative), de la position du mot dans le tronçon et de la position du tronçon dans la phrase.

#### A. Phrase déclarative

Stylisée selon l'hypothèse qu'un certain nombre d'événements mélodiques peuvent être éliminés sans changement perceptif [Vai95], la courbe mélodique est simplifiée sous la forme d'un enchaînement de segments de droite. Les règles du modèle de génération de F0 sont les suivantes :

- La place de l'accent lexical est sauvegardée dans la phrase.
- Au niveau du mot, le contour mélodique se compose de deux segments linéaires : le premier, montant, allant de la première syllabe FD jusqu'à la syllabe accentuée FA (1), et le second, descendant, allant de la syllabe accentuée FD jusqu'à la dernière syllabe FF (2). Le maximum mélodique du mot se réalise sur la syllabe accentuée. Le minimum mélodique se réalise sur la dernière syllabe du mot.
- Au sein d'un tronçon initial, le maximum mélodique des mots augmente du début au premier mot lexical (3) puis diminue jusqu'à la fin du tronçon (4). La fréquence de début se situe autour de la dynamique de base DYB qui correspond à la moyenne des fréquences prélevées sur les syllabes non accentuées du corpus. Elle est estimée à 12 DT. Le maximum mélodique de ce tronçon se réalise sur le premier mot lexical.
- Au sein d'un tronçon intermédiaire, le maximum mélodique des mots augmente du premier au dernier mot du tronçon (5). Le maximum mélodique de ce tronçon se réalise sur le dernier mot du tronçon.
- Au sein d'un tronçon final, (3) et (4). Le maximum mélodique de ce tronçon se réalise sur le premier mot lexical.
- Au sein de la phrase, le maximum mélodique des tronçons diminue du début à la fin de la phrase. Le contour mélodique global est constitué par assemblage des contours de chaque mot. Le minimum mélodique se réalise sur la dernière syllabe de la phrase.
- La ligne haute de déclinaison est la droite qui passe par les maxima mélodiques des tronçons. La ligne de base est la droite qui passe par les minima mélodiques des mots (cf. figure 36).

Exemple :



( wa Da eal ) ( wa la duS Sa GI ru ) ( miH fa Za tu hu ) ( faw qaT TA wi la ti )

**Fig. 36 : Exemple de contour mélodique d'une phrase déclarative.**

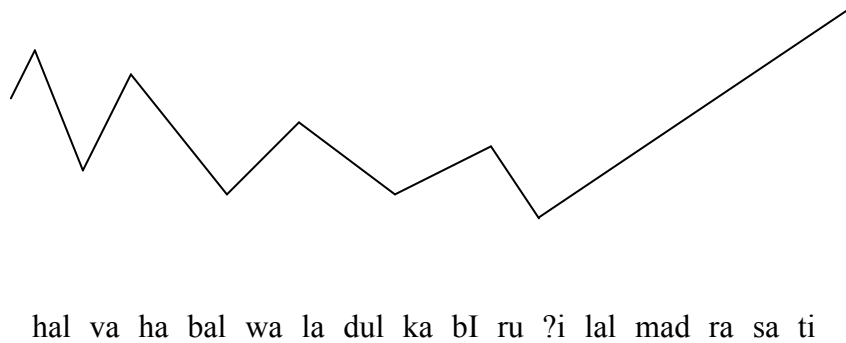
Cette phrase se compose de 4 tronçons : un tronçon verbal, un tronçon nominal sujet, un tronçon nominal objet et un tronçon indirect. Le degré d'accentuation augmente à l'intérieur des tronçons nominal sujet et indirect. Au niveau de la phrase, il diminue du premier au dernier tronçon.

## B. Phrase interrogative

Les règles du modèle de génération du contour mélodique des phrases interrogatives sont les suivantes :

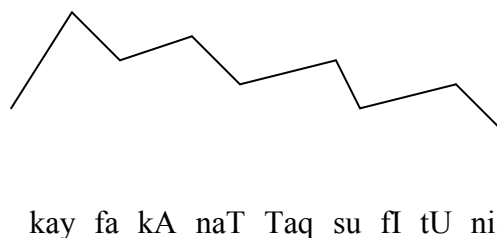
- La place de l'accent lexical est sauvegardée dans la phrase. La dynamique de base DYB est estimée à 12 DT.
- Dans une question totale, le contour mélodique d'un mot non final se compose de deux segments linéaires (1) et (2). Le maximum mélodique de ce mot se réalise sur la syllabe accentuée et le minimum mélodique sur la dernière syllabe. Le contour mélodique d'un mot final se compose d'un segment montant du début à la fin du mot.
- Dans une question partielle, le contour mélodique du mot se compose de deux segments linéaires (1) et (2). Le maximum mélodique du mot se réalise sur la syllabe accentuée et le minimum mélodique sur la dernière syllabe.
- Au sein d'un tronçon initial, le maximum mélodique se réalise sur les marqueurs interrogatifs quand ils existent, sinon sur la première syllabe accentuée.
- Au sein d'un tronçon final d'une question totale, le maximum mélodique se réalise sur la dernière syllabe du tronçon (cf. figure 37).

- Au sein d'un tronçon final d'une question partielle, le maximum mélodique se réalise sur la dernière syllabe accentuée et le minimum mélodique sur la dernière syllabe du tronçon (cf. figure 38).



**Fig. 37 Exemple de contour mélodique d'une question totale.**

- Dans une question partielle, la ligne haute de déclinaison est la droite qui passe par les maxima mélodiques des mots et la ligne de base, la droite qui passe par les minima mélodiques des mots.



**Fig. 38 : Exemple de contour mélodique d'une question partielle.**

Dans ce dernier exemple, la particule /fI/ perd son accent en raison de la collision d'accent avec /tUnis/. Comme nous l'avons déjà mentionné, la succession de deux syllabes accentuées n'est pas tolérée. Celles-ci doivent être séparées par une syllabe inaccentuée ou par une pause [Mar98].

Le modèle que nous avons présenté distingue trois niveaux (mot, tronçon et phrase) pour le calcul des valeurs de F0. Le nombre de syllabes est également pris en compte. Comparé au modèle de Rajouani [Raj89], ce modèle rend compte des phénomènes de déclinaison et opère au niveau des groupes de souffle. De plus, les marqueurs syntaxiques y sont calculés automatiquement. Dans le chapitre 9, nous allons nous concentrer sur

l'évaluation perceptive de l'approche basée sur le tronçon en comparant les phrases synthétiques générées avec ce modèle avec des phrases de recopie de prosodie naturelle.

#### **8.4. Génération de la durée phonémique**

Nous avons développé un modèle à base de règles pour la prédiction automatique de la durée des phonèmes de la langue arabe. Ce modèle étant de type multiplicatif, nous avons procédé en deux étapes : dans un premier temps, les durées intrinsèques de l'ensemble des phonèmes ont été extraites à partir du corpus d'analyse ; dans un second temps, des coefficients de réduction/allongement ont été calculés pour chaque phonème en fonction d'un certain nombre de critères présentés ci-dessous. Les règles sont de la forme :

$$dVoy = CRA * dInt$$

où  $dVoy$  représente la durée du phonème en contexte,  $CRA$  le coefficient de réduction/allongement et  $dInt$  la durée intrinsèque du phonème.

Il existe dans la littérature plusieurs travaux sur la prédiction de la durée en arabe à base de règles [Amr98] [Zem00] ou faisant appel à un réseau de neurones [Chh00]. Ghazali [Gha92a] a présenté une étude sur la relation entre la vitesse d'élocution et la durée des voyelles brèves et longues. Ainsi, il propose que la durée du phonème /a/ soit égale à 91 ms si la vitesse d'élocution est de 4,5 syllabes/seconde et à 54 ms si la vitesse d'élocution est égale à 6,6 syllabes/seconde. Il affirme par ailleurs que la compression des voyelles dans les syllabes fermées est corrélée au débit de la parole.

Zemirli [Zem00] a décrit la relation qui existe entre le nombre de syllabes et la durée des phonèmes dans le mot : plus le mot contient de syllabes, plus la durée des syllabes est réduite sur l'ensemble du mot. Enfin, Amrouche [Amr98] a étudié l'influence des contextes immédiats et la durée des phonèmes. Il a exposé l'impact du voisement sur la durée des phonèmes : l'influence du trait de voisement sur la durée de la voyelle se manifeste par un allongement temporel de cette dernière. Il a également noté que les durées des voyelles brèves et longues étaient différentes selon l'origine des locuteurs.

Nous allons tenter dans ce qui suit de confirmer ou d'infirmer les hypothèses proposées dans la littérature, puis nous proposerons notre modèle de calcul automatique de la durée phonémique.

##### **8.4.1. Les voyelles**

Le tableau 25 représente le nombre, la moyenne et l'écart-type de la durée des voyelles courtes et longues extraites de la base. Nous constatons une différence entre la durée de la voyelle antérieure /a/ et celle des voyelles postérieures /u/ et /i/. Les voyelles longues et leur correspondante brève sont liées par un facteur de proportionnalité  $K$  estimé à 2,5. Le rapport



voyelle longue/voyelle brève a été décrit par Jomaa [Jom94] dans son article sur l'opposition vocalique qui, par ailleurs, présente une compilation de points de vue à ce sujet.

Voyelle	N	Moy	$\sigma$
/a/	1322	93	17,69
/u/	244	87	16,83
/i/	683	82	16,50
/A/	376	227	42,47
/U/	38	218	46,24
/I/	147	213	39,86

**Tab. 25 : Durée des voyelles brèves et longues.**

Les valeurs des voyelles brèves sont inférieures à celles proposées par Mrayati [Mra84] qui situe l'intervalle des voyelles brèves entre 100 et 150 ms et à celles d'El-Ani [Ela70] qui présente des valeurs allant jusqu'à 300 ms. Elles sont par contre proches de celles fournies par Amrouche [Amr98].

- Cas des voyelles pré-pausales

Le tableau 26 représente le nombre, la moyenne, l'écart-type et le coefficient de réduction/allongement CRA de la durée des voyelles courtes précédant une pause. Celles-ci sont plus élevées que les valeurs moyennes précédentes. Rappelons qu'une pause peut se trouver à l'intérieur d'une phrase, clôturant un groupe de souffle, ou en position finale. Ce phénomène n'est pas observé pour les voyelles longues.

Voyelle	N	Moy	$\sigma$	CRA
/a/	94	143	16,23	1,53
/u/	16	140	12,32	1,6
/i/	54	138	17,66	1,68

**Tab. 26 : Durée des voyelles courtes dans une position finale.**

### Règle 1

Les voyelles brèves précédant une pause sont plus longues que les voyelles à l'intérieur des groupes de souffle.

- Cas des syllabes ouvertes/fermées

Le tableau 27 représente le nombre, la moyenne et l'écart-type de la durée des voyelles dans une syllabe ouverte CV et fermée CVC. En ce qui concerne les voyelles longues, seule la voyelle /A/ peut apparaître en syllabe fermée [Gha92a]. Nous constatons que les durées des voyelles dans un contexte CV sont plus élevées que leurs équivalentes dans un contexte CVC. Ce phénomène a été observé dans plusieurs travaux en arabe [Gha92a] [Amr98]. Nous constatons que ceci est plus significatif pour la voyelle longue /A/.

	Syllabe ouverte			Syllabe fermée		
	N	Moy	$\sigma$	N	Moy	$\sigma$
/a/	728	90	18,52	453	84	17,07
/u/	125	91	21,51	96	82	14,56
/i/	311	84	19,68	297	79	16,64
/A/	335	225	41,73	6	190	13,53

**Tab. 27 : Effet de la nature syllabique ouverte/fermée.**

Le tableau 28 représente la durée moyenne des voyelles brèves suivies des phonèmes géminés /ss/ et /bb/. Nous constatons que la compression des voyelles est plus importante lorsqu'elles sont suivies d'un phonème géminé.

	/ss/	CRA	/bb/	CRA
/a/	72	0,77	75	0,8
/u/	75	0,86	78	0,89
/i/	70	0,85	72	0,87

**Tab. 28 : Effet du phonème géminé subséquent.**

## Règle 2

La durée d'une voyelle est moins élevée dans une syllabe fermée que dans une syllabe ouverte. Ceci est d'autant plus vrai si le phonème suivant est géminé.

- Cas des syllabes accentuées

Le tableau 29 représente le nombre, la moyenne et l'écart-type de la durée des voyelles dans une syllabe accentuée et non accentuée. Les voyelles pré-pausales n'ont pas été comptabilisées dans ce calcul.

	Syllabe accentuée			Syllabe non accentuée		
	N	Moy	$\sigma$	N	Moy	$\sigma$
/a/	476	85	17,58	705	88	18,26
/u/	63	86	15,36	158	80	19,30
/i/	182	73	13,53	426	83	19,60
/A/	249	224	40,90	92	227	43,47
/U/	28	218	48,40	5	187	44,18
/I/	83	217	41,91	30	202	38,73

**Tab. 29 : Effet de la nature syllabique accentuée/non accentuée.**

À la lecture de ce tableau, nous constatons qu'il ne semble pas y avoir une corrélation entre la durée des phonèmes et l'accent lexical. En conséquence, le corrélat acoustique de l'accent lexical en arabe est le F0.

### Règle 3

Il n'existe pas de corrélation entre la durée des phonèmes et l'accent lexical.

- Cas du voisement

Le voisement se traduit par un allongement temporel des voyelles. Il a été observé pour le français par Bartkova [Bar87], pour l'anglais par Klatt [Kla76] et pour l'arabe par El-Ani [Ela70] et Amrouche [Amr98]. Nous avons étudié l'influence du voisement à gauche en position V1 dans un contexte C1V1C2V2, avec C1=/s/ voisé et C1=/z/ non voisé. C2 est non voisé (cf. tableau 30)

	C1=/s/			C1=/z/		
	N	Moy	$\sigma$	N	Moy	$\sigma$
/a/	12	85	14,57	6	94	14,8
/u/	6	75	6,48	8	89	11,8
/i/	5	70	3,81	4	86	16,19

**Tab. 30 : Effet du contexte gauche sur la durée des voyelles.**

Ces résultats montrent clairement que la durée des voyelles est plus élevée quand celles-ci sont précédées du phonème /z/ (voisée). Ce phénomène a été noté pour une grande partie des phonèmes voisés. Cependant, nous n'avons observé aucune influence du contexte droit sur la durée des voyelles.

### Règle 4

Les consonnes voisées ont une influence sur la durée des voyelles subséquentes.

- Taille des mots

Nous avons étudié la corrélation entre la taille des mots et la durée des voyelles. La figure 39 (resp. figure 40) représente l'évolution des durées moyennes des voyelles brèves (resp. voyelles longues) en fonction de la taille des mots qui les contiennent exprimée en nombre de syllabes.

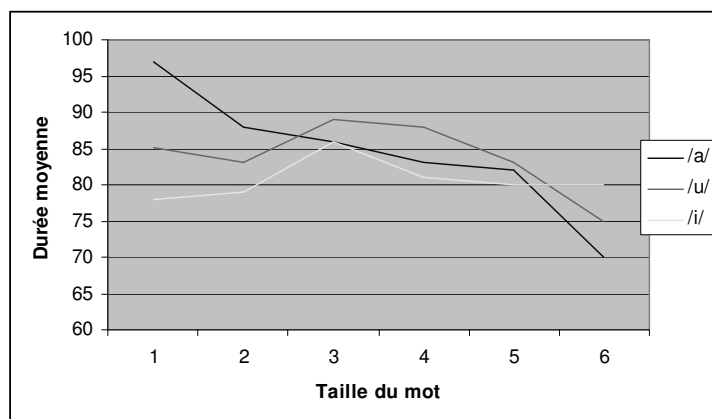


Fig. 39 : Durée des voyelles brèves en fonction de la taille du mot.

Nous remarquons sur ces figures que pour les mots de plus de 3 syllabes, la durée des voyelles diminue au fur et à mesure que le nombre de syllabes augmente. Il semble donc y avoir une relation entre la durée et la taille des mots lorsque celle-ci est supérieure à 3 syllabes.

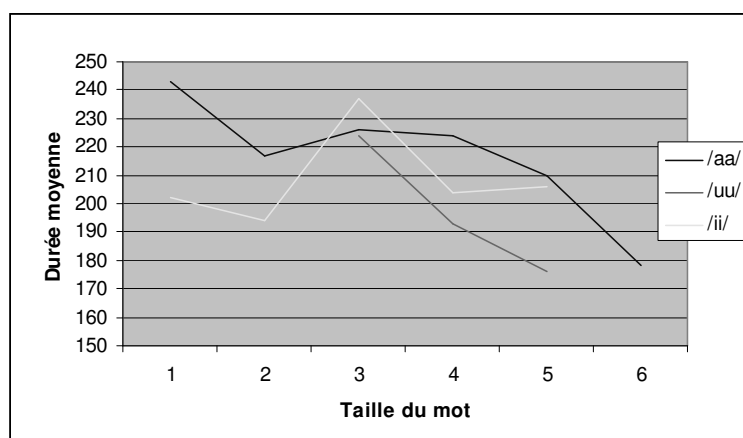


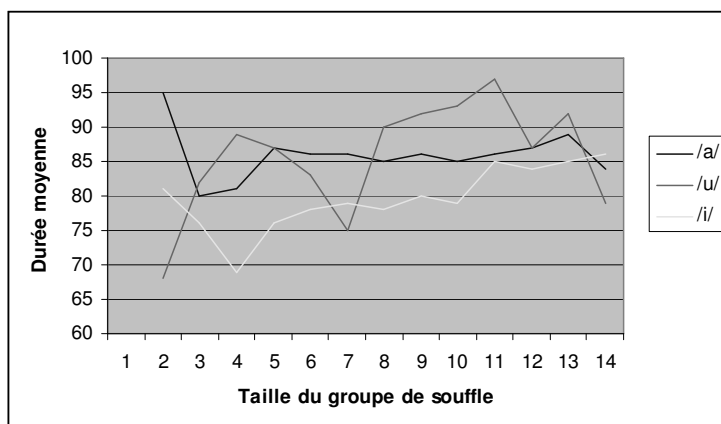
Fig. 40 : Durée des voyelles longues en fonction de la taille du mot.

### Règle 5

Il existe une relation entre la durée des voyelles et la taille des mots de plus de 3 syllabes : plus la taille augmente, plus la durée des voyelles diminue.

- Taille des groupes de souffle

Nous avons étudié la corrélation entre la taille des groupes de souffle et la durée des voyelles. La figure 41 représente l'évolution des durées moyennes des voyelles brèves en fonction de la taille des groupes de souffle exprimée en nombre de syllabes.



**Fig. 41 : Durée des voyelles courtes en fonction de la taille des groupes de souffle.**

En observant cette figure, nous ne pouvons formuler aucune relation entre la durée des voyelles brèves et la taille des groupes de souffle. Ceci a également été observé pour les voyelles longues.

### Règle 6

Il n'existe pas de corrélation entre la durée des voyelles et la taille des groupes de souffle.

#### 8.4.2. Les consonnes

Les travaux sur la prédiction des durées des consonnes arabes sont assez rares. Nous avons extrait dans un premier temps les durées moyennes intrinsèques des consonnes à partir du corpus d'analyse. Le tableau 31 représente les exemples de durées des phonèmes simples /b/, /m/ et /y/.

	N	Moy	$\sigma$
/b/	152	82	18,44
/m/	318	92	21,07
/y/	84	100	20,93

**Tab. 31 : Durée des phonèmes simples /b/, /m/ et /y/.**

#### ○ Cas des phonèmes géminés

Nous avons vérifié sur notre corpus que la durée des phonèmes géminés est approximativement deux fois supérieure à celle des phonèmes simples. Le tableau 32 représente le nombre, la moyenne et l'écart-type de la durée des phonèmes géminés /b/, /m/ et /y/.

	N	Moy	$\sigma$
/bb/	8	176	34,95
/mm/	18	185	34,72
/yy/	56	186	38

**Tab. 32 : Durée des phonèmes géminés /bb/, /mm/ et /yy/.**

### Règle 7

La durée d'un phonème géminé est deux fois supérieure à la durée d'un phonème simple.

- Cas des consonnes non suivies d'une voyelle

Nous avons calculé les durées moyennes des consonnes (dans une position C1) lorsqu'elles sont suivies d'une voyelle d'une part (C1V) et d'une consonne d'autre part (C1C2). Le tableau 33 représente le nombre, la moyenne et l'écart-type de la durée des phonèmes /b/ /d/ /m/ et /w/ dans ces deux contextes.

	Contexte C1V			Contexte C1C2		
	N	Moy	$\sigma$	N	Moy	$\sigma$
/b/	145	80	18,57	23	94	21,27
/d/	135	82	14,69	39	93	16,38
/m/	312	87	16,11	31	96,34	18,80
/w/	222	89	20,88	26	104	14,49

**Tab. 33 : Durée des consonnes dans un contexte C1V et C1C2.**

Nous constatons à partir de ces résultats que la durée des consonnes C1 dans un contexte C1C2 est plus élevée que celle de leur équivalente dans un contexte C1V.

### Règle 8

Les consonnes dont le phonème suivant est une consonne ont une durée supérieure à leur équivalente suivie d'une voyelle.

À l'instar des voyelles, nous n'avons observé aucune relation entre la durée des consonnes et la taille des mots et des groupes de souffle.

### 8.4.3. Présentation du Modèle

Nous proposons la méthode suivante pour le calcul de la durée des phonèmes. CRA est la moyenne de tous les  $CRA_i$  calculés par type de contexte de la manière suivante :

$$CRA = \frac{1}{N} \sum_{i=1}^N CRA_i$$

La formule générale est alors :

$$Durée_{phonème} = CRA \times dIntrinsèque_{phonème}$$

## CHAPITRE 3 : EVALUATION

L'évaluation de la parole synthétique est une étape importante dans le développement d'un système de synthèse vocale. Elle peut être *analytique*, si elle porte sur un des modules appartenant à la chaîne de synthèse, ou *globale*, si elle porte sur la sortie du système. Evidemment, le choix d'un procédé d'évaluation est à considérer en fonction de la méthode de synthèse utilisée : avec une méthode de « stockage/restitution », seule une évaluation globale peut être envisagée, alors que la SAT se prête aux deux approches.

Dans une évaluation analytique, les performances de la transcription orthographique-phonétique, de l'analyse syntaxique, de la prosodie ou du synthétiseur de parole sont mesurées. En général, ces mesures sont effectuées de manière automatique (Ex : aligner la transcription automatique sur la transcription manuelle et compter les erreurs- ce n'est le cas de l'évaluation du synthétiseur de parole qui se fait par des tests subjectifs). Ce sont donc des tests *objectifs* qui obéissent à des règles bien connues. À l'inverse, l'évaluation globale fait intervenir la composante humaine pour juger la performance du système. Elle se base sur des appréciations purement *subjectives* qui n'obéissent à aucune règle formelle.

Pour l'utilisateur naïf d'un système de SAT, c'est la qualité de la parole synthétique globale qui est essentielle. Il existe deux grandes classes de méthodes visant à mesurer cette qualité. La première classe concerne les tests quantitatifs et qualitatifs de l'intelligibilité de la parole. Parmi eux, les tests de rime comme le MRT (Modified Rhyme Test), le DRT (Diagnostic Rhyme Test), le test SUS (Semantically Unpredictable Sentences), etc. [Cal89]. La deuxième classe de méthode a pour but de juger la qualité globale sous forme de questionnaires, soumis à des sujets, ou de scores d'opinion.

Les articles traitant de l'évaluation de la SAT sont nombreux. Citons l'ouvrage « La parole et son traitement automatique » [Cal89] qui présente différentes expériences en laboratoire visant à évaluer la parole codée et synthétique. De leur côté, les articles de [Gol95] et [Gib97] décrivent les critères à prendre en compte lors d'une évaluation d'un système de synthèse vocale. Récemment, BOËFARD et D'ALLESSANDRO ont rédigé un chapitre dans l'ouvrage « Analyse, synthèse et codage de la parole » [Mar02<sup>15</sup>] dans lequel ils présentent un état de l'art des méthodes analytiques et globales.

La dernière étape de notre travail a été l'évaluation de la sortie du système de SAT. Deux tests ont été effectués : un premier test pour l'évaluation de l'intelligibilité et du placement automatique des pauses et un second test pour l'évaluation du contour prosodique. Un protocole d'évaluation a été élaboré dans lequel sont décrits les stimuli utilisés, le choix des sujets et les modalités de leurs réponses. Les résultats obtenus sont décrits ci-dessous.

---

<sup>15</sup> Référence à compléter



## **9.1. Evaluation de l'intelligibilité et du placement des pauses**

### 9.1.1. Présentation du protocole d'évaluation

#### – Les stimuli

Le corpus de textes utilisé est constitué de 21 phrases courtes (3 à 6 mots), moyennes (7 à 11 mots) et longues (plus de 11 mots) pour un total de 198 mots. Les positions des pauses sont prédites automatiquement.

#### – Choix des sujets

Sept sujets ayant une bonne connaissance de la langue arabe ont participé aux tests (4 garçons et 3 filles). Ils n'ont jamais écouté de SAT arabe.

#### – Déroulement du test d'écoute

Toutes les auditions se sont déroulées sous la direction d'un évaluateur. Une première séance de familiarisation (sans évaluation) avec le système complet a été réalisée. Elle a consisté à faire écouter aux auditeurs un ensemble de messages sonores (quelques phrases) en leur fournissant la représentation textuelle correspondante. L'ordre dans lequel les phrases leurs ont été présentées lors de la séance d'évaluation n'est pas important.

#### – Modalité de réponse

Il est demandé aux auditeurs dans cette évaluation de transcrire ce qu'ils écoutent. Il leur est également demandé d'annoter ou de commenter l'insertion ou l'omission des pauses dans les phrases.

### 9.1.2. Résultats

Les sujets n'ont fait aucune remarque en ce qui concerne le placement des pauses. Ceci indique que le modèle de répartition automatique des pauses a été bien accepté. Par contre, ils ont émis des remarques sur certains mots qu'ils ont mal identifiés.

- Les mots suivants ont été mal compris : تنظيم /tanZIm/, أنتقلنا /?intaqalnA/, نظرنا /naZarnA/, أقطار /?aqTAr/, و قد /wa qad/.
- Le mot أبطال /?abTAI/ n'a pas été compris.
- Le mot حاولت /hAwalat/ a été compris هاوالت /hAwalat/.

### 9.1.3. Interprétation

Nous allons tenter d'interpréter ces erreurs de compréhension :

- Certains de ces mots comportent un des phonèmes emphatiques ط/T/, ص/S/, ض/D/, ظ/Z/. Nous rappelons que l'emphase se propage aux phonèmes voisins et que cette propagation n'a pas été modélisée au-delà du contexte immédiat gauche et droit, ce qui a pu altérer la compréhension de ces mots.
- Quatre mots sur sept commencent par la *hamza* /ʔ/ qui a un statut particulier chez les linguistes arabes (cf. chapitre 2). Nous pouvons penser ici, étant donné que la hamza est fortement liée à la voyelle qui la suit, que le problème est lié à la segmentation du diphone /#ʔ/ (# représente le silence).
- L'intelligibilité des phonèmes dépend aussi de la durée qui leur est associée. Une erreur de durée peut altérer la perception analytique des sons, notamment dans le cas des phonèmes géminés (/ʔassasat/).

## 9.2. Evaluation du contour prosodique

### 9.2.1 Protocole d'évaluation

- Les stimuli

Le corpus de textes utilisé est constitué de 30 phrases moyennes (7 à 11 mots) et longues (11 à 18 mots) pour un total de 365 mots. Les phrases naturelles ont été lues à une vitesse moyenne de 10 à 12 phonèmes par seconde. Pour chaque phrase, 5 systèmes sont proposés :

- Un système de voix naturelle (système naturel).
- Un système de voix synthétique avec recopie de prosodie naturelle (système RPN).
- Un système de voix synthétique automatique (système automatique).
- Un système de voix synthétique avec une prosodie plate (système intrinsèque - F0 et durée des phonèmes intrinsèques).

Ainsi, le nombre total des phrases soumises au test est de 120.

- Choix des sujets

Les sujets qui ont participé à ce test sont ceux de la première expérience.

- Déroulement du test d'écoute

Le même évaluateur a dirigé ces auditions. À ce stade, les sujets sont déjà familiarisés avec le système. Dans cette seconde évaluation, les pauses dans les systèmes automatique et intrinsèque ont été placées manuellement en les alignant avec celles du système naturel. Le but étant de neutraliser les incidences d'un mauvais placement automatique sur le jugement des auditeurs. Lors de la séance d'évaluation, les phrases des quatre systèmes ont été présentées dans un ordre aléatoire.

– Modalité de réponse

Les sujets ont attribué une note N à chaque phrase selon l'échelle présentée dans le tableau 34.

Echelle	Qualité
1	Mauvaise
2	Médiocre
3	Passable
4	Bonne
5	Excellente

**Tab. 34 : Echelle d'évaluation.**

### 9.2.2. Résultats

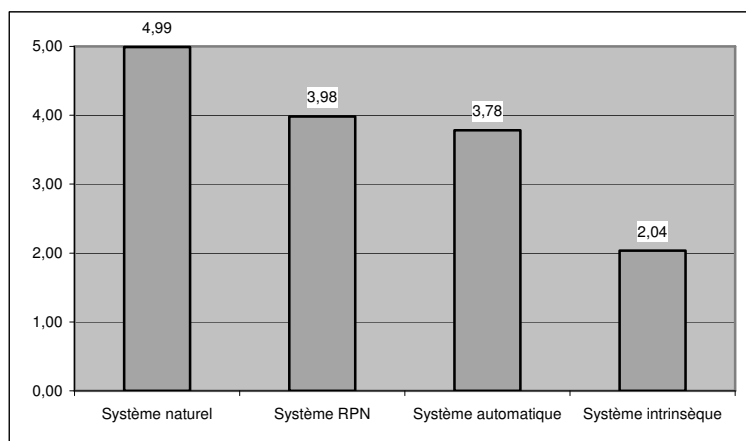
Des notes moyennes d'opinion (Mean Opinion Score ou MOS) ont été calculées pour chaque phrase par type de système (cf. tableau 35). Dans ce tableau, nous pouvons faire les observations suivantes :

- Les phrases du système naturel sont jugées de qualité excellente par l'ensemble des auditeurs.
- Les phrases du système intrinsèque sont jugées de qualité médiocre, à l'exception des phrases 1, 10 et 14 qui sont jugées de qualité passable.
- Les phrases des systèmes RPN et automatique sont jugées de bonne qualité dans leur ensemble.

La figure 42 représente les notes moyennes calculées pour chaque système. Ainsi, l'ordre de préférence des systèmes est le suivant : système naturel (moy = 4,99), système RPN (moy = 3,98), système automatique (moy = 3,78) et système intrinsèque (moy = 2,04). Les écarts sont plus ou moins importants selon les systèmes considérés : le plus grand écart est enregistré entre le système naturel et le système intrinsèque (2,95) et le plus petit entre le système RPN et le système automatique (0,2).

	Système naturel	Système RPN	Système automatique	Système intrinsèque
phrase 1	5,00	4,00	4,00	2,86
phrase 2	5,00	4,00	4,00	2,00
phrase 3	5,00	4,00	3,50	2,00
phrase 4	5,00	4,00	4,00	2,00
phrase 5	5,00	4,00	4,00	2,00
phrase 6	5,00	4,00	4,00	2,00
phrase 7	5,00	3,86	3,57	2,00
phrase 8	5,00	4,00	3,54	2,00
phrase 9	5,00	4,00	4,00	2,00
phrase 10	5,00	4,00	4,00	2,14
phrase 11	5,00	4,00	3,50	2,00
phrase 12	5,00	4,00	4,00	2,00
phrase 13	5,00	4,00	3,46	2,00
phrase 14	5,00	4,00	4,00	2,14
phrase 15	5,00	3,93	3,61	2,00
phrase 16	5,00	4,00	3,50	2,00
phrase 17	5,00	4,00	3,50	2,00
phrase 18	5,00	4,00	3,79	2,00
phrase 19	5,00	4,00	3,43	2,00
phrase 20	5,00	4,00	3,86	2,00
phrase 21	4,86	3,71	3,21	2,00
phrase 22	5,00	4,00	3,50	2,00
phrase 23	5,00	4,00	4,00	2,00
phrase 24	5,00	4,00	4,00	2,00
phrase 25	5,00	3,96	3,50	1,93
phrase 26	5,00	4,00	4,00	2,00
phrase 27	5,00	4,00	4,00	2,00
phrase 28	5,00	4,00	4,00	2,00
phrase 29	5,00	4,00	4,00	2,00
phrase 30	5,00	4,00	4,00	2,00

**Tab. 35 : Notes moyennes d'opinion par phrase.**



**Fig. 42 : Notes moyennes par système.**

La question qui se pose est la suivante : ces différences de moyennes sont-elles significatives, ou bien sont-elles dues au hasard des échantillons ? Pour répondre à cette question, nous avons utilisé le *test d'égalité des moyennes de Student* dont le principe est exposé en Annexe 2.

### 9.2.3. Interprétation des résultats

La **statistique t** de Student calculée pour cet échantillon est comparée à la statistique du modèle  $t_{n-1,p}$  (valeur critique prélevée dans la table de Student au risque  $\alpha$ ). Pour la prise de décision, il faut vérifier si la statistique t appartient au segment formé par les deux valeurs de la variable critique. En d'autres termes, si la statistique t est inférieure à la valeur critique en valeur absolue. Dans ce cas, les moyennes des deux échantillons sont égales. Dans le cas contraire, la différence des moyennes des deux échantillons est significative. Une autre façon de tester si les différences sont significatives est de vérifier si la probabilité du test P est inférieure au risque  $\alpha$ .

Nous avons appliqué le test de Student aux notes moyennes par système avec un risque  $\alpha = 1\%$  et obtenu les résultats présentés dans le tableau 36.

Système1	Système2	Statistique t	Valeur critique	P
Système naturel	Système RPN	145,72	2,75	4,17E-43
Système naturel	Système automatique	26,78		5,25E-22
Système naturel	Système intrinsèque	100,67		1,86E-38
Système RPN	Système automatique	4,66		0,000063
Système RPN	Système intrinsèque	64,27		7,86E-33
Système automatique	Système intrinsèque	35,28		2,24E-25

**Tab. 36 : Résultats du test de Student.**

Nous constatons que la **statistique t** est supérieure à la **valeur critique** (2,75) pour l'ensemble des paires système1-système2, ce qui signifie que les écarts des moyennes obtenus entre les différents systèmes sont significatifs. Nous pouvons donc tirer les premières conclusions suivantes :

- Les phrases du système naturel ont une qualité de contour prosodique supérieure à celle des autres systèmes. Ce résultat, bien qu'attendu, nous a permis de quantifier l'écart de qualité entre le système naturel et les autres systèmes.
- Les phrases du système RPN ont une qualité de contour prosodique supérieure à celle des systèmes automatique et intrinsèque.
- Les phrases du système automatique ont une qualité de contour prosodique supérieure à celle du système intrinsèque.

Pour nous, l'intérêt de ce test est avant tout de situer le système automatique par rapport au système RPN. En considérant le corpus dans son ensemble, le test de Student indique que ces deux systèmes sont différents. À partir de là, nous avons effectué des tests séparés en classant les échantillons de départ selon la taille des phrases correspondantes. L'objectif de ces nouveaux tests est de voir dans quelles conditions la différence des moyennes des systèmes naturel et automatique n'est pas significative.

Deux classes ont été définies : la première classe regroupe les phrases de moins de 12 mots (14 phrases) ; la seconde classe regroupe les phrases de 12 mots et plus (16 phrases). La figure 43 représente les notes moyennes des systèmes RPN et automatique par rapport aux classes 1 et 2. Nous notons que l'écart des moyennes de la classe 1 entre les systèmes RPN et automatique ( $3,99-3,66 = 0,33$ ) est supérieur à celui de la classe 2 ( $3,97-3,89 = 0,08$ ).

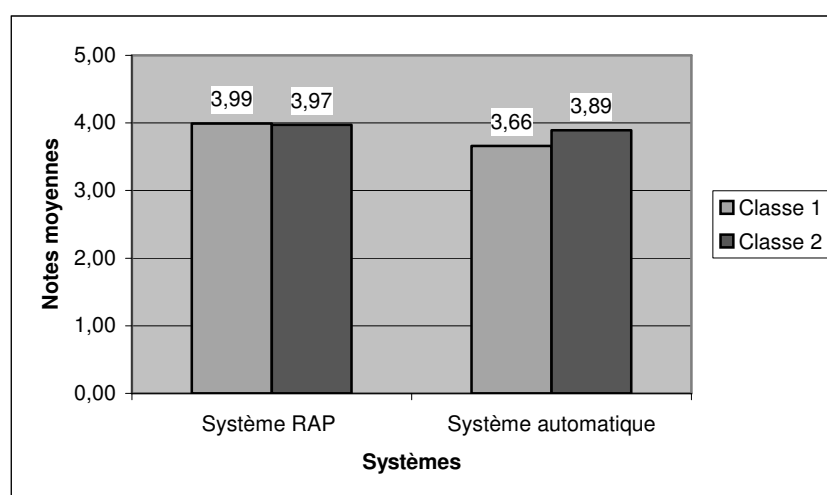


Fig. 43 : Notes moyennes des systèmes RPN et automatique par classe.

Pour vérifier si ces écarts sont significatifs, nous avons appliqué le *test de Student* sur les échantillons des deux classes et obtenu les résultats présentés dans le tableau 37.

	Statistique t	Valeur critique	P
Classe1	5,75	3,01	0,000066
Classe2	1,79	2,94	0,093

**Tab. 37 : Résultats du test de Student en fonction de la taille des phrases.**

En ce qui concerne la classe 1, la statistique t est supérieure à la valeur critique, ce qui signifie que la différence des notes moyennes des systèmes RPN et automatique est significative pour les phrases de moins de 12 mots. À l'inverse, la statistique t est inférieure à la valeur critique dans le cas de la classe 2, ce qui indique que la différence des notes moyennes des systèmes RPN et automatique n'est pas significative pour les phrases de plus de 12 mots. Nous pouvons donc tirer les conclusions suivantes :

- La qualité du contour prosodique des phrases du système RPN est supérieure à celle du système automatique pour les phrases dont la taille ne dépasse pas 12 mots.
- La qualité du contour prosodique des phrases du système RPN est équivalente à celle du système automatique pour les phrases dont la taille est supérieure ou égale à 12 mots.

Les phrases longues ont été préférées aux phrases courtes, ce qui peut paraître paradoxal au premier abord. En effet, Il est bien connu que la charge de concentration des auditeurs est liée à la taille des phrases : plus la phrase est longue, plus la charge de concentration des auditeurs est importante.

Une phrase est une séquence de groupes de souffle séparés par des pauses. Quand elle survient, une marque de pause permet à l'auditeur d'intégrer le message déjà transmis. De ce fait, nous avons reconsidéré le classement précédent sous l'angle des groupes de souffles. Nous avons ainsi classé les phrases de départ selon la taille de leurs groupes de souffle (TGS) et calculé les notes moyennes des systèmes RPN et automatique par rapport à ce nouveau classement (cf. tableau 38).

	Système RPN	Système automatique
TGS < 10	3,9786	3,7643
TGS < 9	3,9813	3,7636
TGS < 8	3,9786	3,7452
TGS < 7	3,9714	3,7179
TGS < 6	4	3,7041

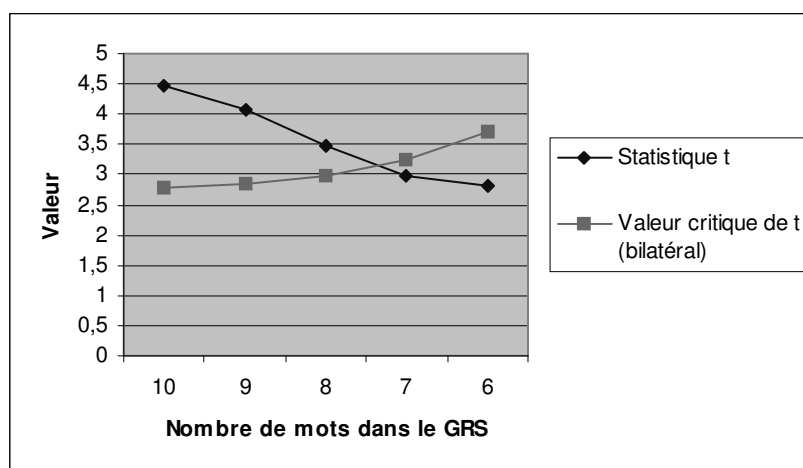
**Tab. 38 : Notes moyennes des systèmes RPN et automatique en fonction de TGS.**

Nous constatons que les notes moyennes du système RPN sont supérieures aux notes moyennes du système automatique. L'application du test de Student, en tenant compte des nouvelles classes, a donné les résultats présentés dans le tableau 39.

	Statistique t	Valeur critique	P
TGS < 10	4,47	2,79	0,00015
TGS < 9	4,07	2,84	0,00059
TGS < 8	3,48	2,97	0,0036
TGS < 7	2,98	3,24	0,0153
TGS < 6	2,81	3,7	0,0304

**Tab. 39 : Résultats du test de Student en fonction de TGS.**

Pour les phrases dont TGS dépasse 7 mots, nous notons que la statistique t est supérieure à la valeur critique. Inversement, la statistique t est inférieure à la valeur critique pour les phrases dont TGS est inférieure à 7 mots. La représentation graphique de ce tableau (cf. figure 44) montre les évolutions de la statistique t et de la valeur critique en fonction de TGS : plus TGS augmente, plus la valeur de la statistique t diminue d'une part et plus la valeur critique augmente d'autre part. Le point de rencontre des deux courbes désigne la valeur de TGS au-dessous de laquelle la différence des notes moyennes des systèmes RPN et automatique n'est plus significative.



**Fig. 44 : Evolution de la statistique t et de la valeur critique en fonction de TGS.**

Ainsi, du point de vue de la phrase, nous avons trouvé que les phrases de grande taille des systèmes RPN et automatique avaient la même qualité, ce qui n'était pas le cas des phrases courtes. Ce résultat, qui est difficile à expliquer, nous a conduits à examiner les



phrases de l'intérieur en considérant la taille du point de vue des groupes de souffle. Nous sommes arrivés au résultat suivant : les phrases des systèmes RPN et automatique qui sont constituées de groupes de souffle de moins de 7 mots sont jugées de même qualité (une phrase peut contenir plusieurs groupes de souffle). L'explication que nous pouvons donner est que la charge mentale des auditeurs serait plus liée à la taille des groupes de souffle, à l'intérieur de la phrase, qu'à la taille globale des phrases. Les pauses constituent donc un indice important pour une bonne perception de la parole.

Notre conclusion est la suivante :

*La qualité du contour prosodique des phrases automatiques est équivalente à la qualité du contour prosodique des phrases avec recopie de prosodie naturelle si la taille de leurs groupes de souffle est inférieure à 7 mots.*

## CONCLUSION GENERALE

L'objectif de notre travail a été de mener une étude dans le domaine de la synthèse de la parole à partir du texte arabe standard voyellé d'une part et l'implémentation des résultats obtenus dans le système multilingue de la société Elan Speech d'autre part. Aujourd'hui, cette société propose la synthèse de la langue arabe à base de diphtonges dans la gamme de ses produits. Le cahier des charges de l'entreprise qui nous a été présenté au départ a été rempli et les objectifs initiaux atteints.

Après avoir présenté le système multilingue d'Elan Speech, qui implémente les technologies de synthèse à base de concaténation de diphtonges (TEMPO™) et à base de corpus (SAYSO™), nous avons décrit le processus de fabrication du dictionnaire de diphtonges. Cette unité présente certaines limites car elle ne couvre pas les phénomènes à long terme comme la propagation de l'emphase. Aussi, les unités dites *sensibles* sont mono-représentées ce qui peut détériorer la qualité du signal synthétique. Néanmoins, le diphtonge reste l'unité qui offre le meilleur rapport qualité/mémoire de stockage : chez Elan Speech, les systèmes de synthèse destinés au monde de l'embarqué (automobile, etc.) ne supportent pas encore la technologie SAYSO en raison des contraintes de stockage.

L'analyse syntaxique est une étape *essentielle* dans un système de synthèse de la parole. Elle doit tenir compte des contraintes de fonctionnement du système (automatique, temps réel, etc.). C'est dans ce contexte que nous avons proposé une grammaire en tronçons qui se fonde sur une analyse superficielle et non-exhaustive du texte. Elle segmente la phrase en groupes de mots non récursifs (intermédiaires entre le mot et la phrase) sans les mettre en relation. L'analyse morpho-syntaxique implémentée repose sur l'utilisation d'un lexique partiel, l'étiquetage par défaut et la propagation de déductions contextuelles.

L'évaluation de l'analyseur morpho-syntaxique a révélé que l'étiquetage par défaut au *cas indirect*, sans l'application des règles contextuelles, donne le meilleur score et que l'étiquetage par défaut au cas direct, le moins bon score. Elle a également démontré que l'introduction des règles contextuelles permet de réduire le taux d'erreurs. Celles-ci ont un impact variable sur les frontières de tronçon : les erreurs les plus critiques sont celles occasionnées par la confusion nom/verbe, les noms propres et les mots se terminant par la lettre *ي/y*. L'amélioration des performances de l'analyseur (3,1% et 4,28% de taux d'erreurs sur les deux corpus) passera sans doute par l'enrichissement des lexiques et des règles contextuelles utilisés.

Après le survol des difficultés en transcription orthographique-phonétique de textes arabes voyellés, nous avons présenté notre système de phonétisation qui repose sur un lexique de mots d'exception et une centaine de règles de réécriture. Pour la modélisation de l'emphase, nous nous sommes limités aux voyelles avoisinant les consonnes emphatiques. La

structure syllabique et la place de l'accent lexical ont été ensuite décrites. Ainsi, nous avons vérifié la validité des règles proposées dans la littérature du point de vue acoustique, notre but étant de reproduire les variations de la fréquence fondamentale. Sous cet angle, nous suggérons que l'accent lexical remonte au-delà de l'antépénultième d'une part et que les syllabes sur-lourdes en fin de mot soient accentuées d'autre part. Nous n'avons pas participé au débat sur l'accent lexical du point de vue de la linguistique.

Nous nous sommes particulièrement intéressés à l'interface syntactico-prosodique qui permet de distribuer les pauses et de générer les paramètres prosodiques de hauteur et de durée. Le modèle de gestion des pauses que nous proposons s'appuie sur les signes de ponctuation, les frontières de tronçons et des seuils phonotactiques qui rendent compte des mécanismes de la phonation. Ainsi, une pause est toujours insérée à un point de ponctuation et peut l'être à une frontière de tronçon si les règles phonotactiques sont vérifiées. Nous avons signifié que ces seules connaissances ne sont pas toujours suffisantes à une bonne prédiction de la place des pauses. Tout d'abord, une pause peut être insérée à la frontière d'un tronçon à l'intérieur d'une phrase imbriquée, étant donné que la mise en relation des tronçons n'est pas effectuée ; ensuite, la distribution des pauses n'est pas déterministe, ce qui rend sa modélisation difficile.

Le contour mélodique global est simplifié sous forme de segments de droite. Il dépend de la modalité de la phrase (déclarative ou interrogative), de la position du mot dans le tronçon et de la position du tronçon dans la phrase. Pour les phrases déclaratives, les analyses statistiques ont démontré que les lignes qui passent par les pics mélodiques des tronçons sont mieux adaptées pour représenter la déclinaison que les lignes passant par les pics mélodiques des mots. Le contour mélodique des phrases interrogatives l'emporte sur les variations à l'intérieur des tronçons. Il dépend du type de marqueur interrogatif quand il existe.

Un modèle multiplicatif à base de règles de réduction/allongement est proposé pour la prédiction des durées phonémiques. Une voyelle est plus allongée dans une position pré-pausale. Elle est, par contre, plus étroite dans une syllabe fermée, en particulier si elle est suivie d'une consonne géminée. Nous n'avons pas noté de corrélation entre la durée des voyelles et l'accent lexical, ce qui nous laisse suggérer que la fréquence fondamentale est le corrélât acoustique le plus pertinent de l'accent lexical. La durée des voyelles est inversement proportionnelle à la taille des mots à partir de 3 syllabes. Ceci n'est pas observé au niveau des groupes de souffle. En ce qui concerne les consonnes, la durée d'une géminée est deux fois supérieure à la durée d'une consonne simple. De plus, la durée d'une consonne dans un contexte non géminé CC est plus élevée que sa durée dans un contexte CV.

L'évaluation du système de synthèse a porté sur la prédiction automatique des pauses et la perception analytique des différents phonèmes, puis sur la qualité du contour prosodique en alignant les pauses avec les stimuli naturels. Les pauses sont bien acceptées par les auditeurs

mais la non modélisation de la propagation de l'emphase altère la compréhension de certains mots. Le contour prosodique des phrases automatiques générées par notre système est jugé équivalent à celui des phrases avec recopie de prosodie naturelle si la taille de leurs groupes de souffle est inférieure à 7 mots. Ce résultat prouve que les pauses sont indispensables à la bonne perception des phrases synthétiques.

Il est certain que l'ensemble de nos conclusions est étroitement lié aux observations faites sur les corpus d'analyse. De la représentativité de ces corpus dépend la robustesse des modèles proposés. Même si nous n'avons pas évalué individuellement les différents composants du système, hormis l'analyseur morpho-syntaxique, nous avons confronté le système à l'utilisateur final et mesuré ses performances en termes d'acceptabilité. L'amélioration de la qualité globale nécessitera avant tout de réduire le taux d'erreurs de chacun des modules et ainsi son incidence sur la sortie du système.

Les perspectives de cette étude sont l'accès à des niveaux supérieurs aux tronçons pour l'étude de la distribution des pauses. Les solutions à envisager doivent suivre la ligne de conduite adoptée dans ce travail sur la souplesse des traitements et la taille des ressources utilisées. Nous avons déjà entamé cette voie en étudiant la méthode d'analyse multilingue de Vergne [Ver02] sur le calcul des relations sujet-verbe. Une autre perspective de ce travail est d'élargir la notion de déclinaison au niveau de la phrase et du paragraphe. En effet, bien que sa pente soit corrélée à la taille des groupes de souffle, certaines expériences ont montré que les extrémités de la ligne de déclinaison sont liées à la position de ces groupes de souffle dans la phrase et à la position de ces phrases dans le paragraphe [Van99a].

Pour terminer, une étude est en cours chez Elan Speech pour développer un système de synthèse à partir du texte arabe à base de corpus. Cette technologie devrait permettre d'améliorer la qualité de la parole synthétique, mais aussi de modéliser certains phénomènes linguistiques, comme la propagation de l'emphase au niveau phonétique.

## BIBLIOGRAPHIE

[Abn91] S. ABNEY, « Parsing by chunks », in R. Berwick, S. Abney, C. Tenny (eds.), Principle-based parsing, Kluwer Academic Publishers, pp: 257-278, Dordrecht.

[Ale01] C. d'Alessandro, E. Tzoukermann, 2001, « Synthèse de la parole », Revue Traitement automatique des langues, Vol. 42:1.

[Aze02] M. AZZEDINE, 2002, « Multilingual Translation System MLTS », International Conference Arabic and Information Technology, Algiers.

[Ali93] M. ALISSALI, 1993, « Architecture logicielle pour la synthèse multilingue de la parole », thèse de doctorat, INPG, GRENOBLE.

[Amr98] A. AMROUCHE, B. BOUDRAA et J.M ROUVAEN, 15-19 juin 1998, « Organisation temporelle des voyelles dans les structures CVCVCV, CVCCVCV et CVCCV de l'arabe standard », Actes des 22<sup>èmes</sup> Journées d'études sur la parole, pp. 91-94.

[Aub91] V. AUBERGE, 1991, « Synthèse de la parole : des règles aux lexiques », thèse de doctorat, Université Pierre Mendès France, Grenoble.

[Bac90] J. BACHENKO & E. Fitzpatrick, 1990, « A Computational Grammar of Discourse-Neutral Prosodic Phrasing in English », Computational Linguistics, Vol. 16, n° 3, pp: 155-170.

[Bal02a] S. BALOUL, P.Y. LE MEUR, 28-29 déc. 2002, « Présentation du système de synthèse de la parole à partir du texte arabe voyellé d'Elan Speech », International Conference Arabic and Information Technology, Algiers.

[Bal02b] S. BALOUL, Ph. BOULA de MAREÜIL, 06-08 mai 2002, « Un modèle syntactico-prosodique pour la synthèse de la parole à partir du texte en arabe standard voyellé », 7ème Conference Maghrebine sur les sciences informatiques, Annaba.

[Bar87] K. BARTKOVA, C. SORIN, 1987, « A model of segmental duration for speech synthesis in French », Speech Communication 6:3, pp. 245-260.

[Bar94] P.A. BARBOSA, « Caractérisation et génération automatique de la structuration rythmique du français », thèse de doctorat, INCP, Grenoble.

[Bea94] F. BEAUGENDRE, « Une étude perceptive de l'intonation du français, développement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte », thèse de doctorat en sciences, Université de Paris XI.

[Bee01] K.R. BEESLEY, 1991, « Finite-state Morphological Analysis and Generation of Arabic », Xerox Research Center, Workshop Arabic Language Processing, pp. 1-8, Toulouse.

[Ben93] B. BENHAMOUDA, 1993, « Les clés de la langue arabe », Office des Publications Universitaires.

- [Ben97] M. BEN HENDA, 1997, « Vers une normalisation des pratiques de communication dans le contexte d'un multilinguisme intégral (arabe-latin) : Entre la diversité des usages et les contraintes technologiques », rapport de recherche, CEM-GRESIC, MSHA Université de Bordeaux 3.
- [Big93] D. Bigorgne, O. Boëffard, B. Cherbonnel, F. Emerard, D. Larreur, J.L. Le Saint-Milon, I. Metayer, C.Sorin & S. White , 1993, « Multilingual Psola text-to-speech system », ICASSP, Minneapolis, vol. 2, pp. 187-190.
- [Big93] D. BIGORGNE, 1993, « Interface de programmation des modules SYC V1 », Note technique, France Télécom R&D.
- [Bla75] R. BLACHERE, 1975, « Grammaire de l'arabe classique », Éditions Maisonneuve / Larousse, Paris.
- [Boh79] G. BOHAS, 1979, « Contribution à l'étude de la méthode des grammairiens arabes en morphologie et en phonologie d'après les grammairiens arabes tardifs », thèse de doctorat, Université de Lille 3.
- [Boi00] R. BOITE, H. BOURLARD, T. DUTOIT, J. HANCQ & H. LEICH, 2000, « Traitement de la parole », chapitre « Synthèse de la parole à partir d'un texte », pp. 345-441, Collection Electricité, Presses polytechniques et universitaires romandes.
- [Bou97] P. BOULA DE MAREUIL, 1997, « Etude linguistique appliquée à la synthèse de la parole à partir du texte », thèse de doctorat, Université de Paris XI, Orsay.
- [Bou01] P. BOULA DE MAREÜIL, P. CELERIER, T. CESSSES, S. FABRE, C. JOBIN, P.-Y. LE MEUR, D. OBADIA, B. SOULAGE, J. TOËN, 2001, « Elan Text-To-Speech : un système multilingue de synthèse de la parole à partir du texte », Traitement Automatique des Langues, Vol. 42, n° 1, pp: 223-252.
- [Cal89] CALLIOPE, 1989, « La parole et son traitement automatique », édition MASSON, Paris.
- [Cam92] W. CAMPBELL, 1992, « Syllable-based segmental duration », Editions G. Bailly et C. Benoît, Talking Machines: Theories, Models and Designs, Elsevier Science Publishers, pp. 211-224.
- [Cam98] E. CAMPIONE & J. VERONIS, 1998, « A multilingual prosodic database », ICSLP, pp: 3163-3166, Sydney, Australie.
- [Cam01] E. CAMPIONE, 2001, « Etiquetage prosodique semi-automatique des corpus oraux », TALN, Tours.
- [Can60] J. CANTINEAU, 1960, « Etude de la linguistique arabe », Librairie C. Klincksieck, Paris.
- [Chh00] A. CHEHAB, A. ZAKI, A. RAJOUANI, 28-30 juin 2000, « Un modèle neuronal pour la prédiction de la durée des syllabes de la langue arabe », Actes des 23<sup>èmes</sup> Journées d'études sur la parole, Aussois, pp. 97-100.

[Che00] N. CHENFOUR, A. BENABBOU, A. MOURADI, 2000, « Etude et Evaluation de la Di-Syllabe comme Unite Acoustique pour le Système de Synthèse Arabe PARADIS », 2<sup>nd</sup> International Conference on Language Resources & Evaluation », Athens, Greece.

[Con99] A. Conkie, 1999, « Robust unit selection system for speech synthesis », Acoustical Society of America meeting, Berlin.

[Coo00] G. Coorman, J. Fackrell, P. Rutten & B. Van Coile, 2000, « Segment selection in the L&H RealSpeak laboratory TTS system », ICSLP, Beijing.

[Deb98] F. DEBILI, H. ACHOUR, 1998, « Voyellation automatique de l'arabe », ACL '98, Montréal.

[Dej98] H. Dejean, 1998, « Découverte de structures syntaxiques à partir de corpus », thèse de doctorat de l'Université de Caen.

[Dut96a] T. DUTOIT, 1996, « An Introduction to Text-To-Speech Synthesis », Kluwer Academic Publishers.

[Dut96b] T. DUTOIT, V. PAGEL, N. PIERRET, O. VAN DER VREKEN, F. BATAILLE, 1996, « The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes », *Proc. ICSLP 96*, Philadelphia.

[Ela70] M. EL-ANI, 1970, « Arabic phonology: An acoustical and physiological investigation », Mouton, The Hague, Paris.

[Elg01] A.M. ELGENDY, « Aspects of Pharynged Coarticulation », PhD thesis, University of Amsterdam.

[Elk90] J. EL KAFI, 16 février 1990, « Contribution à la réalisation d'un système multilingue de synthèse de la parole à partir de texte autour d'un processeur spécialisé : Le TMS50C42 », thèse de doctorat, Université de Bordeaux I.

[Eme77] F. EMERARD, 1977, « Synthèse par diphtonges et traitement de la prosodie », thèse de Doctorat, Université de Grenoble III.

[Ess88] ES-SKALI, 7 mars 1988, « Eléments d'un modèle intonatif pour la synthèse de la parole arabe », thèse de doctorat, Université MOHAMED V, Faculté des Sciences de RABAT.

[Fer02] K. FERRAT, 2002, « Apport des transitions formantiques pour la synthèse par règles de la parole en arabe standard », International Conference Arabic and Information Technology, Alger, Algérie, pp. 14-24.

[Fuj67] H. FUJISAKI, 1967, « Physics of speech sounds », Tokyo, University Press.

[Gau98] A. GAUDINAT et J.P. GOLDMAN, 15-19 juin 1998, « Le système de synthèse FIPSVOX : syntaxe, phonétisation et prosodie », Actes des 22<sup>èmes</sup> journées d'études sur la parole, Martigny, pp. 139-142.

- [Gha77] S. GHAZALI, 1977, « Back consonants and backing coarticulation in Arabic », Ph.D. thesis, Université de Texas.
- [Gha81] S. GHAZALI, 1981, « La diffusion de l'emphase : l'inadéquation d'une solution tauto-syllabique », Analyse théorie.
- [Gha92a] S. GHAZALI, A. BRAHIM, 19-24 mai 1992, « Voyelles longues et voyelles brèves en arabe standard : Organisation temporelle », Actes des 19<sup>èmes</sup> Journées d'études sur la parole, Bruxelles, pp. 153-154.
- [Gha92b] S. GHAZALI, M.ZRIGUI, Z. MILED et H. JEMNI, 19-22 mai 1992, « Synthèse de l'arabe standard à partir du texte par TD-PSOLA : Le traitement des processus phonologiques », Actes des 19<sup>èmes</sup> Journées d'études sur la parole, Bruxelles, pp. 89-93.
- [Gib97] D. GIBBON, R. MOORE, R. WINSKI, 1997, « Handbook of Standards and Ressources for Spoken Language Systems », Editions Moutons de Gruyter, Berlin & New York.
- [Gig98] E. Giguet, 1998, « Méthode pour l'analyse automatique de structures formelles sur documents multilingues », thèse de doctorat de l'Université de Caen.
- [Gol95] M. GOLDSTEIN, 1995, « Classification of methods used for assessment of text-to-speech systems according to the demands placed on the listener », Speech Communication 16, pp. 225-244.
- [Gue83] M. GUERTI, 1983, « Contribution à la synthèse de la parole en arabe standard » (synthèse par diphtones et technique de prédiction linéaire), thèse de magister, Université d'Alger.
- [Gue87] M. GUERTI, octobre 1987, « Contribution à la synthèse de la parole en arabe standard », Actes des 16<sup>èmes</sup> Journées d'études sur la parole, Hammamet, pp. 290-292.
- [Had79] A. HADJ-SALAH, 1979, « Linguistique arabe et linguistique générale », Essai de méthodologie et d'épistémologie du ilm al-Arabiyya, 2 vol.
- [Ham00] W. M. HAMZA et A. RASHWAN, 2000, « Concatenative Arabic Speech Synthesis Using Large Speech Database », ICSLP, Beijing, China.
- [Han00] A.N. HANNA et N.A. GHATTAS, Mars 2000, « Text-To-Speech Synthesis of Arabic », Workshop on Friendly Exchanging Through the Internet, ENSERB, Bordeaux, France.
- [Har92] M. HARKAT, 1992, « Phonétique et phonologie », Editions Dar El Afak, Alger.
- [Jol01] J. M. JOLION, 2001, « Probabilités et Statistique », support de cours, Département Génie Productique, INSA de Lyon.
- [Jom94] M. JOMAA, 1-3 juin 1994, « L'opposition de durée vocalique en Arabe : Essai de typologie », Actes des 20<sup>èmes</sup> Journées d'études sur la parole, Trégastel, pp. 395-400.



- [Kel92] E. KELLER, 1992, « Le choix d'un modèle informatique pour la prosodie en synthèse de la parole », rapport de recherche, Université de Lausanne.
- [Kha99] A. EL-KHAIRY, 1999, « The Arabic Pharyngeal Approximant », ICPhs99, San Francisco, pp. 1029-1032.
- [Kla76] D.H. KLATT, 1976, « Linguistic uses of segmental duration in English : acoustic and perception evidence », Journal of the Acoustical Society of America, 59, pp. 1208-1221.
- [Kla80] D.H. KLATT, 1980, « Software for a cascade/parallel synthesizer », Journal of the Acoustical Society of America, 67 (3), pp. 971-995.
- [Kou76] D. KOULOUGHLI, 1976, « Contribution à l'étude de l'accent en arabe littéraire », Annales de l'Université d'Abidjan, série H, vol. IX, pp. 124-125.
- [Lem96] P.Y LE MEUR, 1996, « Synthèse de la parole par unités de taille variable », thèse de doctorat, ENST, Télécom Paris.
- [Lib92] M.Y. LIBERMAN & C.W. CHURCH, 2002, « Text Analysis and Word Pronunciation in Text-to-Speech Synthesis », in S. FURUI & M.M. SONDHAI (eds.), Advances in Signal Processing, pp: 791-831, Dekker, New York.
- [Mar02] MARIANI, 2002,
- [Mer00] P. MERTENS, J.P. GOLDMAN, E. WEHRLI et A. GAUDINAT, 2000, « La synthèse de l'intonation à partir de structures syntaxiques riches », soumis pour une publication, 37 pp.
- [Mou84] A. MOURADI, A. RAJOUANI et M. NAJIM, 28-30 mai 1984, « La synthèse de l'arabe à partir du texte », Actes des 13<sup>èmes</sup> Journées d'études sur la parole, Bruxelles, pp. 153-154.
- [Mou87] A. MOURADI, 1987, « Validité et limites du diphone en tant qu'unité de synthèse pour la langue arabe », Actes des 16<sup>èmes</sup> Journées d'études sur la parole, Hammamet, pp. 296-297.
- [Mou90] E. MOULINES, F. CHARPENTIER, 1990, « Pitch-Synchronous Waveform Processing Techniques for Text-To-Speech Synthesis Using Diphones », Speech Communication, Vol 9, pp. 453-467.
- [Mor97] Y. MORLEC, 1997, « Génération multi-paramétrique de la prosodie du français par apprentissage automatique », thèse de doctorat, Institut de la Communication Parlée.
- [Mor98] M. MOREL et A. LACHERET- DUJOUR, 15-19 juin 1998, « Utilisation d'une structure arborescente pour une hiérarchisation fine des règles de transcription graphème-phonème », Actes des 22<sup>èmes</sup> Journées d'études sur la parole, Martigny, pp. 151-154.
- [Mra84] M. MRAYATI et J. MAKHOUL, « Man-Machine Communication and the Arabic Language », Rapport technique.

- [Naj98] Z. NAJIM, 15-19 juin 1998, « Contour intonatif et syntaxe en arabe : Résultats préliminaires », Actes des 22<sup>èmes</sup> Journées d'études sur la parole, Martigny, pp. 155-158.
- [Ohm67] S. OHMAN, 1967, « World and sentence intonation : a quantitative model », Quarterly Progress and Status Report, 2, K.T.H., Stockolm, pp. 20-54.
- [Osh81] D. O'SHAUGHNESSY, 1981, « A study of French vowel and consonant durations », Journal of Phonetics, pp. 385-406.
- [Oue02] R. OUERSIGHNI, 2002, « A major offshoot of the DIINAR-MBC project : AraParse, a morpho-syntactic analyzer for unvowelled Arabic texts », Workshop Arabic Language Processing, Toulouse, pp. 9-16.
- [Pas90] V. PASDELOUP, 1990, « Modèles de règles rythmiques du français appliqué à la synthèse de la parole », thèse de doctorat, Institut de Phonétique d'Aix-en-Provence, Université de Provence, Aix-Marseille I.
- [Pie81] J. PIERREHUMBERT, 1981, « Synthesizing intonation », Journal of the Acoustical Society of America, 70, Editions R.B. Lindsay, New York, pp. 985-995.
- [Que92] H. QUENÉ & R. KAGER, « The derivation of prosody for text-to-speech from prosodic sentence structure », Computer Speech and Language, Vol. 6, n° 1, pp: 77-98.
- [Raj89] A. RAJOUANI, 1989, « Contribution à la réalisation d'un système de synthèse à partir du texte pour l'arabe », thèse de doctorat, Université Mohamed V, Faculté des Sciences de RABAT.
- [Ros71] M. ROSSI, 1971, « Le seuil de glissando ou seuil de perception des variations tonales », Phonética 23, pp. 1-33.
- [Rut00] P. RUTTEN, G. COORMAN, J. FACKRELL & B. VAN COILE, 2000, « Issues in corpus based speech synthesis », Seminar IEE, State of the art in speech synthesis, Savoy Place, pp. 16/1-16/7.
- [Sab89] G. SABAH, 1989, « L'intelligence artificielle et le langage : processus de compréhension », vol. 2, Hermès, Paris.
- [Saf01] N.D. SAFA, A.N HANNA & A. RAJOUANI, 2001, « Enhancement of a TTS System for Arabic Concatenative Synthesis by Introducing a Prosodic Model », ACL-EACL Workshop on Arabic Language Processing, pp: 97-102, Toulouse, France.
- [Sag90] Y. SAGISAKA, 1990, « On the prediction of global F0 shapes for Japanese text-to-speech » IEEE; International Conference on Acoustics, Speech and Signal Processing 1: pp. 325-328.
- [Sar90] A. SAROH, J. BRUSSET et J. TIHONI, 1990, « Vers une production automatique de textes phonétiques pour l'arabe standard à partir de sa graphie », Actes des 18<sup>èmes</sup> Journées d'études sur la parole, Montréal, pp. 305-309.

- [Sch02] J. L. SCHWARTZ, P. ESCUDIER, 2000, « La parole : des modèles cognitifs aux machines communicantes », chapitre « la synthèse de la parole », pp. 213-245, Hermès science publications.
- [Sor95] C. SORIN et F. EMERARD, 1995, « Domaines d'application et évaluation de la synthèse de parole à partir du texte », rapport interne, CNET, Lannion.
- [Sty96] Y. STYLIANOU, 1996, « Modèles harmoniques plus bruit combinés avec des méthodes statistiques pour la transformation de la parole et du locuteur », thèse de doctorat, ENST (Télécom Paris).
- [Tai97] N. TAIBI, 1997, « Contribution à l'étude du traitement automatique des erreurs dans un texte écrit en arabe », thèse de magister, Alger.
- [Tra92], C. TRABE, 1992, « F0 generation with a database of natural F0 patterns and with a neuronal network », Editions G. Bailly et C. Benoît, Talking Machines: Theories, Models and Designs, Elsevier Science Publishers B.V., pp. 287-304.
- [Vai95] J. VAISSIÈRE, « Phonetic Explanations for Cross-Linguistic Prosodic Similarities », *Phonetica*, Vol. 52, pp: 123-130.
- [Van99a] G. VANNIER, 1999, « Etude des contributions des structures textuelles et syntaxiques pour la prosodie: application à un système de synthèse à partir du texte », thèse de doctorat, UFR de Sciences, Université de Caen.
- [Van99b] G. Vannier, A. Lacheret-Dujour et J. Vergne, 1999, « Pauses location and duration calculated with syntactic dependencies and Textual considerations for T.T.S. SYSTEM », XIVth International Congress of Phonetic Sciences (ICPhS), San-Francisco, California.
- [Ver98a] J. Vergne & E. Giguet, 1998, « Regards théoriques sur le tagging » », TALN, pp: 22-31, Paris, France.
- [Ver98b] J. Vergne, 1998, « Entre arbre de dépendance et ordre linéaire, les deux processus de transformation : linéarisation et reconstruction de l'arbre », cahiers de Grammaire n° 23, Toulouse, France, pp. 95-136.
- [Ver99] J. Vergne, 1999, Étude et modélisation de la syntaxe des langues à l'aide de l'ordinateur. Analyse syntaxique automatique non combinatoire. Habilitation à Diriger des Recherches, Université de Caen.
- [Ver02] J. VERGNE, 24-27 juin 2002, « Une méthode pour l'analyse descendante et calculatoire de corpus multilingues : application au calcul des relations sujet-verbe », TALN 2002, Nancy.
- [Yvo96] F. YVON, 1996, « Prononcer par analogie : motivation, formalisme et évaluation », thèse de doctorat, ENST, Paris.
- [Yvo98] F. YVON, P. BOULA DE MAREUIL, C D'ALESSANDRO, V. AUBERGE, M. BAGEIN, G. BAILLY, F. BECHET, S. FOUKIA, J-F. GOLDMAN E. KELLER, D. O'SHAUGHNESSY, V. PAGEL, F. SANNIER, J. VÉRONIS et B. ZELLNER, 1998,

« Objective Evaluation of Grapheme to Phoneme Conversion for Text-To-Speech Synthesis in French », *Computer Speech and Language*, pp. 393-410.

[Zak00a] A. ZAKI, A. RAJOUANI, 2000, « Synthesis of arabic speech from text : Towards a system of high quality », *Workshop on Friendly Exchanging Through the Internet*, ENSERB, Bordeaux, France, pp. 45-48.

[Zak00b] A. ZAKI, A. RAJOUANI, M. NAJIM, 2000, « Contours intonatifs de la phrase interrogative en arabe », 23<sup>èmes</sup> Journées d'études sur la parole, Aussois.

[Zem98a] Z. ZEMIRLI, 1998, « SYNTHAR+ : Synthèse vocale sous MULTIVOX », *Technique et Science informatique*, 17(6).

[Zem98b] Z. ZEMIRLI, N. VIGOUROUX, A. HENNI et G. PERENNOU, 10-12 juin 1998, « Quel modèle morphologique, lexical et phonologique pour le traitement automatique de la langue arabe ? », *TALN*, Paris.

[Zem00] Z. ZEMIRLI, N. VIGOUROUX, 28-30 juin 2000, « Vers une modélisation de la durée des sons pour la génération automatique du rythme dans la synthèse de la langue arabe », *Actes des 23<sup>èmes</sup> Journées d'études sur la parole*, Aussois, pp. 261-264.

## Liste des tableaux

Tab. 1 : Applications du système d'Elan Speech. ....	15
Tab. 2 : Exemple d'analyse morphologique du mot /ktb/. ....	25
Tab. 3 : Exemples de voyellation en fonction du sens et du contexte. ....	26
Tab. 4 : Normes de codage ASMO et leurs équivalentes ISO. ....	27
Tab. 5 : Les flexions casuelles de la langue arabe. ....	44
Tab. 6 : Formes dérivées à l'accompli de la racine trilitère /kasb/. ....	45
Tab. 7 : Exemple de génération du verbe /ʔittaxava/. ....	46
Tab. 8 : Liste des étiquettes verbales. ....	53
Tab. 9 : Liste des étiquettes nominales. ....	54
Tab. 10 : Liste des étiquettes au cas indirect. ....	55
Tab. 11 : Table de compatibilité antéfixe-préfixe. ....	58
Tab. 12 : Matrices de compatibilité. ....	62
Tab. 13 : Matrice de confusion du corpus 1. ....	63
Tab. 14 : Résultats de l'étiquetage par défaut aux différents cas. ....	64
Tab. 15 : Matrice de confusion du corpus 2. ....	64
Tab. 16 : Répartition des erreurs d'étiquetage du corpus 1. ....	65
Tab. 17 : Répartition des erreurs d'étiquetage du corpus 2. ....	65
Tab. 18 : Répartition des erreurs. ....	67
Tab. 19 : Pré-traitements des textes dans le système de SAT d'Elan Speech. ....	75
Tab. 20 : Distribution des pauses. ....	103
Tab. 21 : Indicateurs de surface. ....	104
Tab. 22 : Distribution des pauses non associées aux signes de ponctuation. ....	105
Tab. 23 : Distribution des pauses par type de tronçon. ....	107
Tab. 24 : Pentas de déclinaison en fonction de la taille des phrases. ....	124
Tab. 25 : Durée des voyelles brèves et longues. ....	129
Tab. 26 : Durée des voyelles courtes dans une position finale. ....	129
Tab. 27 : Effet de la nature syllabique ouverte/fermée. ....	130
Tab. 28 : Effet du phonème géminé subséquent. ....	130
Tab. 29 : Effet de la nature syllabique accentuée/non accentuée. ....	130
Tab. 30 : Effet du contexte gauche sur la durée des voyelles. ....	131
Tab. 31 : Durée des phonèmes simples /b/, /m/ et /y/. ....	133
Tab. 32 : Durée des phonèmes géminés /bb/, /mm/ et /yy/. ....	134
Tab. 33 : Durée des consonnes dans un contexte C1V et C1C2. ....	134
Tab. 34 : Echelle d'évaluation. ....	139
Tab. 35 : Notes moyennes d'opinion par phrase. ....	140
Tab. 36 : Résultats du test de Student. ....	141
Tab. 37 : Résultats du test de Student en fonction de la taille des phrases. ....	143
Tab. 38 : Notes moyennes des systèmes RPN et automatique. ....	143
Tab. 39 : Résultats du test de Student. ....	144

## Liste des figures

Fig. 1 : Schéma général d'un système de synthèse à partir du texte.....	11
Fig. 2 : Architecture du système de SAT d'Elan Speech.....	17
Fig. 3 : Exemple d'une phrase voyellée /yavhabUna limuddati sanatin/.....	20
Fig. 4 : Exemple de voyellation de la phrase.....	25
Fig. 5 : Interface de l'outil Sonalog.....	37
Fig. 6 : Analyse syntaxique traditionnelle.....	39
Fig. 7 : Analyse syntaxique à base de <i>tagger</i> .....	40
Fig. 8 : Mécanisme de dérivation en arabe.....	42
Fig. 9 : Exemple de dérivation du mot /ktb/.....	42
Fig. 10 : Structuration du lexique arabe.....	43
Fig. 11 : Exemple de découpage en tronçons (entre parenthèses).....	48
Fig. 12 : Diagramme bloc de l'analyse linguistique.....	52
Fig. 13 : Analyse morphologique.....	60
Fig. 14 : Module de transcription orthographique-phonétique.....	77
Fig. 15 : Système syllabique de la langue arabe.....	82
Fig. 16 : Représentations fréquentielle et temporelle de la phrase /?addarsul eAOir/.....	89
Fig. 17 : Vue générale de Prosel.....	90
Fig. 18 : Evolution de F0 pour la phrase /waqaea naZruhu ealA namlatin/.....	94
Fig. 19 : Représentation du registre du locuteur.....	94
Fig. 20 : Exemple de réinitialisation de la courbe intonative.....	95
Fig. 21 : Modèles intonatifs des phrases arabes.....	98
Fig. 22 : Contour intonatif au niveau du mot.....	99
Fig. 23 : Exemple d'un contour intonatif pour une phrase en arabe.....	99
Fig. 24 : Règles du modèle intonatif MIR.....	100
Fig. 25 : Interface graphique d'ElanStudio.....	102
Fig. 26 : Distribution de la taille des petits groupes de souffle.....	106
Fig. 27 : Hiérarchisation des durées des pauses.....	109
Fig. 28 : Exemple d'insertion d'une pause après la mise en relation des tronçons.....	110
Fig. 29 : Evolution de F0 en fonction de la position.....	118
Fig. 30 : Evolution de F0 en fonction de la position des syllabes.....	118
Fig. 31 : Pente moyenne en fonction.....	119
Fig. 32 : Coefficient de corrélation en fonction.....	120
Fig. 33 : Pente moyenne en fonction.....	121
Fig. 34 : Coefficient de corrélation en fonction.....	121
Fig. 35 : Comparaison des coefficients de corrélation des approches 1 et 2.....	122
Fig. 36 : Exemple de contour mélodique d'une phrase déclarative.....	126
Fig. 37 Exemple de contour mélodique d'une question totale.....	127
Fig. 38 : Exemple de contour mélodique d'une question partielle.....	127
Fig. 39 : Durée des voyelles brèves en fonction de la taille du mot.....	132
Fig. 40 : Durée des voyelles longues en fonction de la taille du mot.....	132
Fig. 41 : Durée des voyelles courtes en fonction de la taille.....	133
Fig. 42 : Notes moyennes par système.....	141
Fig. 43 : Notes moyennes des systèmes RPN et automatique par classe.....	142
Fig. 44 : Evolution de la statistique t et de la valeur critique.....	144

## Annexe 1

Les symboles utilisés dans ce document et leurs équivalents en IPA

Symbole arabe	Symbole Elan Speech	IPA
ء ئ و ا إ	?	?
ب	b	b
ت ة	t	t
ث	c	θ
ج	j	ʒ
ح	H	ħ
خ	x	x
د	d	d
ذ	v	ð
ر	r	r
ز	z	z
س	s	s
ش	O	š
ص	S	ʂ
ض	D	ɖ
ط	T	ʈ
ظ	Z	ʈʂ
ع	ε	ʕ
غ	G	ɡ
ف	f	f
ق	q	q
ك	k	k
ل	l	l
م	m	m
ن	n	n
ه	h	h
و	w	w
ي	y	y
اَ	a	a
اُ	u	u
اِ	i	i
اَ	A	a:
اُ	U	u:
اِ	I	i:

## Annexe 2

### Principe du test de Student [Jol01]

Soient  $X_1$  et  $X_2$  deux lois normales de moyennes  $\mu_1$  et  $\mu_2$ , et d'écart types  $\sigma_1$  et  $\sigma_2$ . Le test repose sur les hypothèses suivantes :

- **Hypothèse nulle  $H_0$**  :  $\mu_1 = \mu_2$ , c'est-à-dire que les moyennes des deux échantillons sont égales.
- **Hypothèse alternative  $H_1$**  :  $\mu_1 \neq \mu_2$ , c'est à dire que les moyennes des deux échantillons sont différentes.

On dispose de deux échantillons de tailles  $n_1$  et  $n_2$  sur lesquels on peut faire des estimations de moyennes  $m_1$  et  $m_2$  et d'écart types  $s_1$  et  $s_2$ .

#### Calcul de la statistique T

- Si les écart types  $\sigma_1$  et  $\sigma_2$  sont connus :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

L'hypothèse **H0** au risque  $\alpha$  est rejetée si  $z \notin \left[-t_1 - \frac{\alpha}{2}, t_1 - \frac{\alpha}{2}\right]$  où la valeur  $t_1 - \frac{\alpha}{2}$  est lue dans la table de la loi normale centrée réduite.

- Si les écarts-types  $\sigma_1$  et  $\sigma_2$  sont inconnus, alors il faut tenir compte de la taille des échantillons

a) Si  $n_1$  et  $n_2$  sont tous les deux supérieurs à 30 :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

L'hypothèse **H0** au risque  $\alpha$  est rejetée si  $z \notin \left[-t_1 - \frac{\alpha}{2}, t_1 - \frac{\alpha}{2}\right]$  où la valeur  $t_1 - \frac{\alpha}{2}$  est lue dans la table de la loi normale centrée réduite.

b) Si  $n_1$  ou  $n_2$  est inférieur à 30 et  $\sigma_1 = \sigma_2$ :

$$z = \frac{m_1 - m_2}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



où

$$\hat{\sigma} = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}$$

L'hypothèse **H0** au risque  $\alpha$  est rejetée si  $z \notin \left[ -t_1 - \frac{\alpha}{2}; n_1 + n_2 - 2, t_1 - \frac{\alpha}{2}; n_1 + n_2 - 2 \right]$  où la

valeur  $t_1 - \frac{\alpha}{2}; n_1 + n_2 - 2$  est lue dans la table de Student à  $n_1 + n_2 - 2$  degrés de liberté.

c) Si  $n_1$  ou  $n_2$  est inférieur à 30 et  $\sigma_1 \neq \sigma_2$ :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

L'hypothèse **H0** au risque  $\alpha$  est rejetée si  $z \notin \left[ -t_1 - \frac{\alpha}{2}; v, t_1 - \frac{\alpha}{2}; v \right]$  où la valeur critique

$t_1 - \frac{\alpha}{2}; v$  est lue dans la table de Student à  $v$  degrés de liberté;  $v$  étant l'entier le plus proche de

$$\frac{\left[ \frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1} \right]^2}{\frac{s_1^4}{(n_1 - 1)^3} + \frac{s_2^4}{(n_2 - 1)^3}}$$