Interface syntaxe-prosodie dans un système de synthèse de la parole à partir du texte en arabe

S. Baloul[1,3], *M. Alissali*[1], *M. Baudry*[1] *& P. Boula de Mareüil*[2,3]

[1] Laboratoire d'Informatique de l'Université du Maine
F-72085 Le Mans CEDEX 9
Tél. : +33 (0) 2 43 83 38 74
Fax : +33 (0) 2 43 83 38 68
baloul@lium.univ-lemans.fr

[2] LIMSI-CNRS
BP 133
F-91403 Orsay CEDEX
Tél. : +33 (0) 1 69 85 81 19
Fax : +33 (0) 1 69 85 80 88
mareuil@limsi.fr

[3] Elan Informatique
4 rue Jean Rodier
F-31400 Toulouse
Tél. : ++33 (0) 5 61 36 07 77
Fax : ++33 (0) 5 61 36 07 70
http://www.elan.fr

ABSTRACT

This paper presents a syntactico-prosodic model and its implementation in a diphone Arabic text-to-speech (TTS) system. This model, based on rewrite rules, first calculates the syntactic markers of the input text. Second, a phrasing operation segments it into chunks. The syntax-prosody interface then enables the allocation of pauses and the generation of prosodic parameters: the melodic contour depends on the sentence modality, on the word position within chunks and on the chunk position within the sentence. The implemented modules are currently being evaluated within a multilingual TTS system assessment.

1. INTRODUCTION

La contribution de la syntaxe à la prosodie a été mise en évidence dans les systèmes de synthèse traitant de diverses langues indo-européennes [Bou01]. En ce qui concerne la langue arabe, peu de recherches ont porté sur 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 [Raj89] : elles supposent une analyse syntaxique sophistiquée et un générateur prosodique fondé sur la structure syntaxique ainsi produite. Mais à défaut d'analyse syntaxique automatique pour l'arabe, l'étiquetage des mots se fait manuellement, ce qui est inenvisageable dans un but de système automatique de synthèse de la parole.

Des recherches plus récentes réfutent cette nécessité et suggèrent que la prosodie peut être générée indépendamment, sur la base de critères acoustiques, phonologiques et phonotactiques [Saf01]. La démarche proposée ici se distingue de ces deux tendances et prône une position intermédiaire : pour nous, la syntaxe est incontournable, mais une analyse syntaxique superficielle, partielle (*shallow/partial parsing*) peut suffire au calcul de la prosodie. Ce traitement syntaxique est entièrement automatisable, au moins lorsque le texte est voyellé.

La section suivante pose les principes d'une grammaire en tronçons et présente dans ses grands traits la méthode suivie, apprise sur un corpus (A) de quelque 200 phrases traduites issues de MULTEXT [Cam98], adaptées à l'arabe et voyellées par un expert, et d'un corpus (B) également voyellé, constitué de 120 phrases isolées, dont les structures syntaxiques varient progressivement (des structures simples aux structures complexes), et dont la longueur varie de 2 à 8 mots, dans la lignée de [Raj89]. Ce second corpus a été lu par un locuteur algérois de 28 ans, ayant une bonne maîtrise de l'arabe standard, et analysé grâce à un outil de recopie de prosodie développé à Elan [Bou01].

Comme illustré (figure 1), le découpage des phrases en tronçons nécessite la connaissance de la catégorie grammaticale des mots qui les composent : c'est le rôle de l'analyse morpho-syntaxique (section 2.2). Le parenthésage syntaxique (section 3), qui est également à base de règles, est exploité pour prédire le placement de l'accent lexical et des pauses, et pour calculer l'évolution de la fréquence fondamentale (F_0).

2. LA GRAMMAIRE EN TRONÇONS : APPLICATION A L'ARABE

2.1 Principes généraux

Notre objectif ici est de produire une analyse syntaxique en vue de la synthèse vocale. L'analyse syntaxique n'est donc pas un but en soi, mais 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.

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. La pertinence de cette unité dans la hiérarchie mot-tronçon-phrase a été démontrée dans différentes langues [Gig98].

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 : d'une certaine manière, il peut être comparé au groupe accentuel (ou *mot prosodique*) [Ver98]. La question que nous nous sommes dès lors posée est la suivante : comment délimiter ces tronçons en arabe ? Nous allons tenter d'y apporter un début de réponse en section 3.

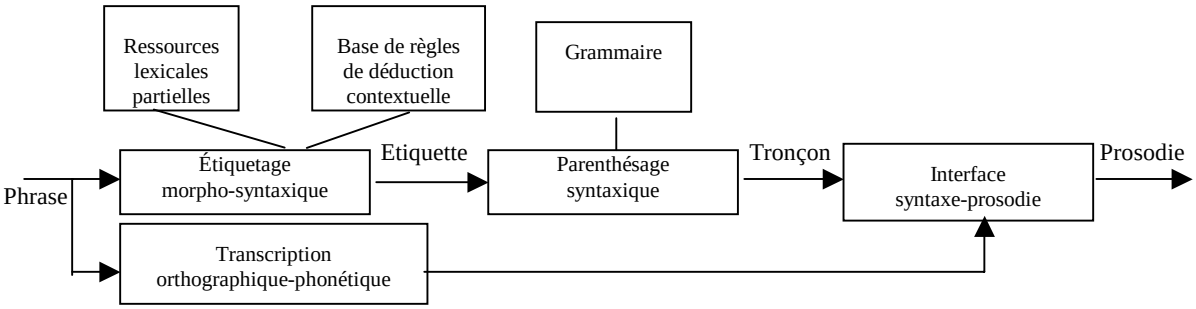

Figure1 : Diagramme bloc de l'analyse linguistique.

Mais en préalable aux prises de décisions sur les dépendances entre les mots, il faut définir un jeu d'étiquettes grammaticales adapté au découpage en tronçons.

2.2 Étiquetage morpho-syntaxique

L'étiquetage morpho-syntaxique (ou *tagging*) que nous avons développé reprend les principes de Vergne [Ver98]. L'approche repose sur la propagation de déductions contextuelles utilisant un lexique partiel : une étiquette par défaut étant associée à chaque mot, les règles de déductions contextuelles interviennent en aval pour confirmer la valeur attribuée par défaut, ou au contraire, modifier cette valeur en fonction du contexte d'apparition du mot.

Morphologie arabe : Dans ce qui nous intéresse ici, nous privilégions les mécanismes purement morphologiques, sans référence à la position du mot dans la phrase. Selon la grammaire traditionnelle, le lexique arabe comprend trois catégories de mots : verbes, noms (substantifs et adjectifs) et particules (adverbes, conjonctions et prépositions). Hormis les noms propres, les mots des deux premières catégories sont dérivés à partir d'une racine : un squelette de trois consonnes radicales le plus souvent. À partir d'une racine, passée dans différents *schèmes*, une famille de mots peut être engendrée autour d'un même concept sémantique : c'est le fait le plus caractéristique de la morphologie arabe. Un autre fait remarquable est le caractère flexionnel des mots : les terminaisons permettent de distinguer le mode des verbes et la fonction des noms. Ainsi, la *damma*, la *fatha et* la *kasra* sont des indices très importants pour nos règles.

Analyse morphologique : 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.

Une difficulté de l'arabe en traitement automatique est l'agglutination par laquelle les composantes du mot sont liées les unes aux autres. Ainsi notre étiqueteur morpho-syntaxique identifie-t-il d'abord les composantes du mot. Nous avons adopté la segmentation de Zemirli [Zem98], d'où le découpage pour le mot يَسْأَلُونَهَا. Nous avons à cet effet élaboré des tables de compatibilité entre les différents éléments du mot.

L'ordre du traitement est le suivant : particules, verbes et noms. Pour le traitement des verbes, à côté des 14 formes connues dans la littérature, nous avons défini 14 autres formes pour les verbes malades. En plus de ce lexique de schèmes verbaux, nous utilisons des lexiques partiels de mots grammaticaux (particules), de déclinaisons nominales, de préfixes et de suffixes. Ces lexiques sont enrichis de mots spécifiques.

Désambiguïsation : Les déductions contextuelles sont exprimées dans le même formalisme que l'analyse morphologique, à travers une vingtaine de règles. Les règles sont locales — elles agissent sur un mot et ses proches voisins : leur portée est de 2 à 3 mots maximum. Ces règles doivent être ordonnées : par exemple, la règle qui réécrit l'étiquette X en Y doit intervenir avant les règles appelant comme contexte une étiquette Y, X et Y étant des étiquettes quelconques.

L'ensemble des règles de segmentation et de désambiguïsation est réalisé à l'aide de *flex*, un générateur d'analyseur lexical, qui implémente une forme étendue des expressions régulières.

3. PARENTHÉSAGE SYNTAXIQUE

Après l'étiquetage morpho-syntaxique des mots, 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, qui détermine la structure de la phrase *verbale* et dont la structure est répertoriée sous la forme de schèmes prédéfinis ainsi que d'éventuelles lettres additionnelles ;
- l'accord entre les unités, qui a trait notamment aux variations en genre (masculin ou féminin) et en nombre (singulier, duel ou pluriel) ;
- l'ordre des unités, dont certaines unités non récursives comme les couples *déterminé + déterminant* et *nom + épithète* se combinent selon un ordre fixe. 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 désambiguïsation contextuelle.

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 défini et à des contraintes d'accords fortes, est assimilée à un tronçon. À partir de là, nous avons défini quatre types de tronçons :

1- tronçon verbal : regroupant un verbe et d'éventuelles particules le précédant (de négation, interrogative...). Exemple : مَعَهُ (لَمْ يَذْهَبْ) ، (أَكَلَ) الْخُبْزَ ;

2- tronçon sujet : pouvant être introduit par des particules et regroupant les formes *nom sujet + complément du nom* et *nom sujet + épithète*. Exemple :

(الرَّجُلُ الْعَجُوزُ) (مَرِيضٌ) ;

3- tronçon objet : pouvant être introduit par des particules et regroupant les formes *nom objet + complément du nom* et *nom objet + épithète*. Exemple :

إِنَّ (فَرِيقَ الْعَاصِمَةِ) قَوِيٌّ ;

4- tronçon indirect : regroupant les formes *prépositions + complément indirect*, la tête restant nominale. Exemple : (بَعْدَ الأَكْلِ) (إِلَى الْمَدْرَسَةِ) ذَهَبَ ;

Pour déterminer les séquences d'étiquettes appartenant à un même tronçon, nous avons défini une relation de compatibilité : *si deux étiquettes successives sont compatibles, alors elles appartiennent au même tronçon*. Cette relation est exprimée dans des matrices (tableau 1) 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 classes correspondant aux : sujet, objet, indirect, verbe et particule. Quant à la conjonction de coordination, elle a un statut particulier dans la mesure où les constituants qui l'entourent sont regroupés au sein d'un même tronçon si, et seulement si, ils ont la même étiquette.

4. INTERFACE SYNTAXE-PROSODIE

La sortie de l'analyse syntaxique, qui fournit un alignement de mots et d'étiquettes grammaticales ainsi que la suite de tronçons, est connectée aux modules suivants de mise en correspondance prosodique. Une frontière mineure est associée à la fin des tronçons (#fm), une frontière majeure après un signe de ponctuation faible (#FM) et une frontière terminale en fin de phrase (#FT). Exemple :

يُمكِّنَا (#fm) أَن نُبدِيَ (#fm) فَرحَةً غَامِرَةً ، (#FM) أَو حُزنًا . (#FT)

Les unités ainsi délimitées ne constituent pas des groupes de souffle séparés par des pauses. Celles-ci sont ajustées avec le nombre de syllabes et gérées par le module phonotactique — la phonétisation et la syllabation ne sont pas décrites ici.

Tableau 1 : Exemple de matrice de compatibilité de la classe sujet (Nsi, Nsd, Nsa, Nss désignent les noms sujets respectivement indéterminés, déterminés par l'article, par annexion et par suffixation ; X désigne les étiquettes autres que celles de la ligne ; 0 indique que les étiquettes peuvent apparaître au sein d'un même tronçon, 1 que non ou que la suite n'est pas attestée en arabe).

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

4.1 Gestion des pauses

Générer des pauses est indispensable à l'intelligibilité de la parole synthétique. Pour ce faire, nous nous sommes appuyés sur les signes de ponctuation et nous avons défini des seuils minimal et maximal de syllabes non séparées par une pause, estimés à 8 et 14 respectivement, qui rendent compte de contraintes physiologiques pesant sur la phonation et la respiration. Une pause est toujours insérée aux frontières #FM et #FT, et peut l'être à une frontière #fm si le nombre de syllabes depuis la dernière pause est supérieur au seuil minimum, le nombre de syllabes jusqu'à une frontière #FM ou #FT suivante est supérieur à 4 syllabes et que l'une des conditions suivantes est vérifiée :

- le nombre de syllabes depuis la dernière pause est supérieur au seuil de 14 syllabes ;
- le tronçon suivant est de type indirect (c'est-à-dire, introduit par une préposition) ;
- le tronçon suivant commence par une conjonction de coordination ;
- la frontière sépare des tronçons objet (ou indirect) et verbe.

Exemple : أَوَدُّ أَنْ أَطْلُبَ سَيَّارَةَ أُجْرَةٍ تَأْتِي غَدًا صَبَاحًا فِي سَاعَةٍ مُبَكِّرَةٍ
pause

4.2 Placement de l'accent lexical

Diverses études ont mis en relation groupe syntaxique et groupe accentuel [Bou97]. L'accent est le phénomène de mise en relief de certaines syllabes qui sont perçues de manière plus forte que les syllabes voisines. En arabe, les études en prosodie considèrent traditionnellement l'existence de deux niveaux d'accent, en plus du niveau inaccentué : l'accent primaire et l'accent secondaire. Leur position est prédictible : elle dépend du nombre et des types de syllabes contenus dans le mot. Les règles appliquées ici pour la prédiction automatique de la position de l'accent sont définies dans [Ela70].

4.3 Réalisation du contour mélodique

Nous appelons *accent de tronçon* l'accent porté par le dernier mot du tronçon. Théoriquement, sur la courbe de

F_0 d'un mot isolé arabe, le maximum de fréquence fondamentale se situe sur la syllabe qui porte l'accent primaire [Raj89]. L'analyse perceptive du corpus B nous amène à dire que, dans la phrase, tout mot arabe garde son accent lexical, à l'exception des monosyllabes entrant en collision syllabique avec le mot qui les suit. Le degré d'accentuation des mots augmente au fur et à mesure qu'on se rapproche de la fin du tronçon. En même temps et inversement, l'accent de tronçon diminue au fur et à mesure qu'on se rapproche de la fin de la phrase. Ce phénomène de déclinaison peut être représenté par deux lignes : une ligne haute et une ligne de base.

Le taux de déclinaison est fonction de la longueur de la phrase — par exemple, elle décroît par pas de demi-tons pour une phrase de 10 syllabes. Des remises à zéro peuvent également intervenir quand une pause est insérée, si le nombre de syllabes est suffisant. L'exemple ci-dessous (figure 2) illustre l'augmentation du degré d'accentuation à l'intérieur des tronçons sujet et indirect, et sa diminution progressive au niveau de la phrase.

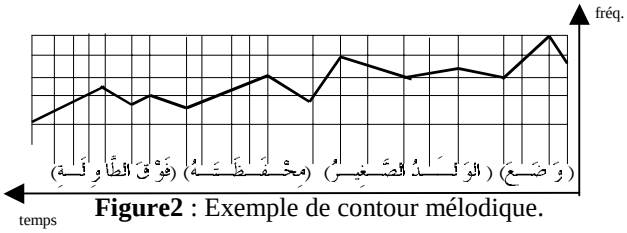

Figure2 : Exemple de contour mélodique.

Stylisée selon l'hypothèse qu'un certain nombre d'événements mélodiques peuvent être éliminés sans changement perceptif [Tha91], la courbe mélodique est simplifiée sous la forme d'un enchaînement de segments de droite. En sortie, chaque phonème est caractérisé par ses hauteurs de départ et d'arrivée, ainsi que par sa durée (tableau 2).

Tableau 2 : Exemple de représentation prosodique pour le mot ذهب (/ðahaba/)— le registre pour une voix d'homme est <90 Hz ; 135 Hz>. La hauteur initiale de chaque phonème est raccordée à la hauteur finale du phonème précédent, si celle-ci est non nulle.

(ذَهَبَ) (الوَلَدُ) (إِلَى المَدْرَسَةِ)(بَعْدَ الأَكْلِ)						
	ð	a	h	a	b	a
hauteur initiale Hz)	110	121	0	129	123	120
hauteur finale (Hz)	121	134	0	123	120	114
durée (ms)	92	106	80	95	50	95

Pour les durées, un modèle multiplicatif a été implémenté : des facteurs d'allongement/ réduction sont appliqués aux durées intrinsèques des phonèmes.

5. CONCLUSION

Nous avons présenté dans cet article un nouveau modèle syntactico-prosodique pour l'arabe standard voyellé, intégré dans le système multilingue de synthèse de la parole à partir du texte de la société Elan Informatique [Bou01]. Ce système est en cours d'évaluation depuis la phonétisation jusqu'à la qualité globale du système.

L'évaluation du module d'analyse morpho-syntaxique sur un nouveau corpus de 200 phrases a donné un taux d'erreur de 8 % sur les étiquettes entraînant 3% d'erreurs sur les frontières de tronçons. Nous avons recensé les sources d'erreurs les plus importantes pour notre étiqueteur, dont l'impact sur les frontières de tronçons est variable : erreurs de segmentation de mots dont les éléments de base sont pris pour des préfixes ou des suffixes ; erreurs non corrigées par le contexte sur des noms dont la structure ressemble à celle de verbes ; erreurs sur des verbes malades qui ne sont pas reconnus comme verbes.

La prochaine étape de ce travail consiste à valider, par des tests d'écoute, les résultats actuels et à refaire l'expérimentation sur un corpus peut-être plus riche, tant au niveau de la longueur des phrases que de la diversité des structures syntaxiques, afin d'affiner l'analyse. L'un des résultats escompté est la validation de nos tables de compatibilité morphologiques et syntaxiques.

BIBLIOGRAPHIE

[Abn91] Abney S. (1991), « Parsing by chunks », *Principle-based parsing*, Kluwer Academic Publishers, Dordrecht, pp. 257-278.

[Boh79] Bohas G. (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 de l'université de Lille III.

[Bou97] Boula de Mareüil P. (1997), *Étude linguistique appliquée à la synthèse de la parole à partir du texte*, thèse de doctorat de l'Université de Paris XI, Orsay.

[Bou01] Boula de Mareüil P. *et al* (2001), « Elan Text-To-Speech : un système multilingue de synthèse de la parole à partir du texte », Traitement Automatique des Langues **42**(1), pp. 223-252.

[Cam98] Campione E. & Veronis J. (1998), « A multilingual prosodic database », *ICSLP*, Sydney, pp. 3163-3166.

[Ela70] El-ani M. (1970), *Arabic phonology: An acoustical and physiological investigation*, Mouton & Co., The Hague, Paris.

[Gig98] Giguet E. (1998), *Méthode pour l'analyse automatique de structures formelles sur documents multilingues*, thèse de doctorat de l'université de Caen.

[Raj89] Rajouani A. (1989), *Contribution à la réalisation d'un système de synthèse à partir du texte pour l'arabe*, thèse de doctorat de l'université Mohammed V, Rabat.

[Saf01] Safa N. D. *et al.* (2001), « Enhancement of a TTS System for Arabic Concatenative Synthesis by Introducing a Prosodic Model », *ACL-EACL Workshop on Arabic Language Processing*, Toulouse, pp. 97-102.

[Tha91] 't Hart J. *et al.* (1991), *A perceptual study of intonation: an experimental-phonetic approach to speech method*, University Press, Cambridge.

[Ver98] Vergne J. & Giguet E. (1998), « Regards théoriques sur le "tagging" », *TALN*, Paris, pp. 22-31.

[Zem98] Zemirli Z. (1998), « SYNTHAR+ : Synthèse vocale sous MULTIVOX », *Technique et science informatique* **17**(6).