

Reconnaissance de la parole en Arabe Moderne : le système MARS

Mahieddine Djoudi

Résumé : Nous présentons dans cet article le système MARS (Modern Arabic Recognition System) de reconnaissance automatique de la parole continue en Arabe moderne. Le système a pour but la compréhension de phrases dans un contexte multilocuteur. Sa mise en œuvre tient compte des particularités de la langue aux niveaux phonétique et linguistique. Nous décrirons, dans un premier temps le système de décodage acoustico-phonétique [9], ensuite, nous aborderons le traitement linguistique en évoquant les aspects de la morphologie de la langue, et de la structure syntaxico-sémantique des phrases. Nous parlerons, enfin, de l'importance de la prosodie dans un système de compréhension de l'Arabe parlé.

Mots clés : Arabe moderne, Parole continue, Décodage

1 Introduction

L'Arabe moderne a fait l'objet de plusieurs travaux récents ayant trait, soit à l'aspect phonétique [2] [11] soit à la composante linguistique de la langue [3], [5]. Toutefois, le problème de la reconnaissance automatique n'a été que très peu abordé jusqu'à présent [4] [7]. La reconnaissance de la parole continue fait appel à diverses sources de connaissances. Cette démarche conduit à la conception de systèmes complexes ayant de fortes interactions entre les différents modules qui les composent. Le système que nous proposons a pour but la compréhension de phrases en arabe moderne, dans un contexte multilocuteur. Sa mise en œuvre tient compte des particularités de la langue.

2 Architecture du système

La structure générale du système fait apparaître deux grands sous-systèmes. D'une part, un système de décodage acoustico-phonétique, en l'occurrence SAPHA, d'autre part, le décodeur linguistique SALAM. Chaque sous système utilise diverses sources de connaissances. Le décodeur phonétique fournit au système linguistique, en mode proposition, un treillis phonétique. Il peut être réactivé, en mode vérification, par le décodeur linguis-

tique pour savoir si un phonème donné est l'étiquette d'un morceau de signal donné. Le décodeur linguistique part des données fournies par le système phonétique et de celles provenant de la définition du langage, détermine la signification du message oral (voir figure ??).

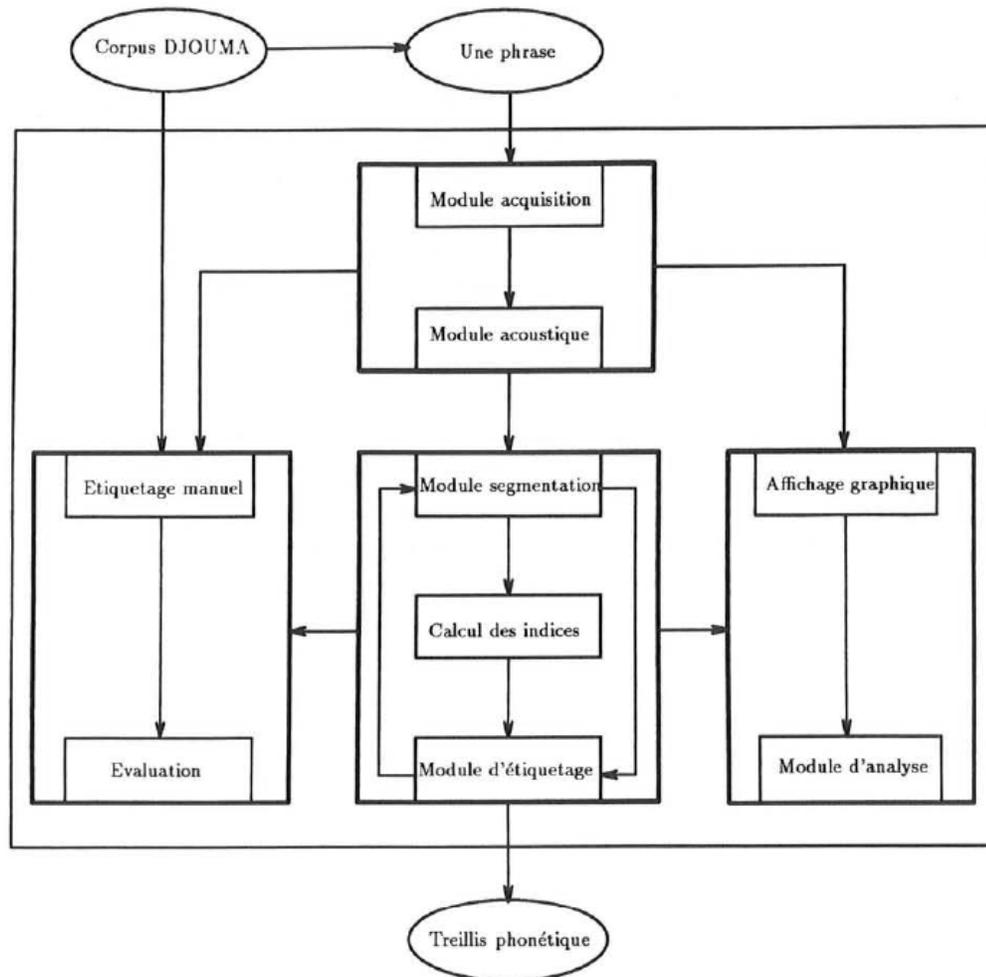


FIGURE 1 – Architecture de Mars

3 Décodeur acoustico-phonétique

Ce système est considéré comme le premier maillon du processus de compréhension. Il construit la description phonétique d'un énoncé, à partir du signal acoustique numérisé. Les principales étapes du système sont :

- la segmentation du signal en grandes classes phonétiques en utilisant des algorithmes non contextuels et reposant sur des critères simples. Les classes retenues sont les voyelles, les plosives, les fricatives et les sonnantes.
- le calcul des indices phonétiques pertinents à la reconnaissance. Ces indices sont : la durée, le degré de voisement, la barre d’explosion, le suivi de formants, les transitions formantiques, le centre de gravité et la limite du bruit de friction.
- l’identification phonétique des segments en utilisant un système expert à base de règles de production qui affecte des étiquettes phonétiques aux segments obtenus lors de la segmentation en utilisant les indices phonétiques et les connaissances se trouvant dans la base de règles.

Le résultat du système est un treillis de phonèmes, représenté par une suite de segments. A chaque segment, est associée une liste des phonèmes les plus probablement prononcés par le locuteur. L’évaluation du système, qui consiste en la comparaison du décodage automatique avec l’étiquetage manuel de 50 phrases phonétiquement équilibrées et prononcées par 3 locuteurs masculins donne actuellement un score global de 65%.

4 Analyseur linguistique

SALAM comme Système d’Analyse Linguistique de l’Arabe Moderne reçoit en entrée une suite de phonèmes sous forme de treillis phonétique (résultats fournis par le décodeur SAPHA) et effectue la reconnaissance de la phrase traitée. Son résultat consiste pour l’instant en l’interprétation sémantique et l’écriture de la phrase prononcée. Le décodeur linguistique SALAM comporte plusieurs modules :

4.1 Module morphologique

L’objectif de ce module est de détecter des mots dans le treillis de phonèmes fourni en entrée, de vérifier leur appartenance à la langue et de fournir les valeurs grammaticales fixes (catégorie syntaxique, genre, nombre, temps, mode, cas, ...) Le processus peut être déclenché de différentes manières :

- **la vérification de mots** : cela se produit lorsque le module syntaxico-sémantique émet une hypothèse concernant la présence d’un mot particulier à un instant donné de la phrase. Cette hypothèse, qui constitue une donnée du module morphologique, est un couple composé :
 - d’une forme de référence représentant une transcription phonétique du mot-hypothèse

- et d'un repère temporel indiquant le début du mot dans le treillis de phonèmes. Le traitement consiste à apparier la forme de référence et le sous-treillis correspondant au mot à vérifier. Le résultat est une validation ou un rejet de l'hypothèse traitée.
- **la recherche de mots** : lorsque le module ne dispose pas d'information temporelle sur un mot à valider, il procède à une recherche itérative en effectuant un balayage du treillis de phonèmes. A chaque itération, un appariement, entre le schème de référence et la partie sélectionnée dans le treillis, est tenté. Cette opération séquentielle rend le processus très coûteux en temps de traitement. Pour réduire ce dernier, on peut s'appuyer sur les hypothèses prosodiques qui vont permettre d'éliminer les appariements avec un sous-treillis contenant un marqueur prosodique.

Difficultés de l'analyse morphologique

Les problèmes rencontrés pendant l'analyse morphologique peuvent être de deux types, les uns sont liés à la structure radicale de la langue et les autres aux erreurs introduites par le décodage acoustico-phonétique.

Les difficultés liées à la langue : Les problèmes spécifiques à la langue arabe les plus importants sont :

- N'importe quelle racine ne peut pas être instanciée à n'importe quel schème.
- Deux racines analogues ne donnent pas lieu à des dérivations analogues.
- Cas de la racine faible, c'est à dire comportant les consonnes /w/ ou /y/.
- Cas des mots homographes, ou une même chaîne recouvre deux notions différentes suivant le contexte, exemple /fi/ verbe impératif de /wafa/ et /fii/ préposition.

Les incertitudes du décodage phonétique : En l'état actuel du décodeur phonétique, le treillis phonétique contient différents types d'erreurs :

- Une sur-segmentation peut provoquer la duplication de certains phonèmes.
- Une sous-segmentation risque de faire disparaître d'autres phonèmes.
- D'autre part, pour un phonème prononcé, le système SAPHA fournit plusieurs solutions possibles.
- De plus, il arrive que le phonème réellement prononcé n'apparaisse pas dans le choix proposé.

Pour résoudre les problèmes de sur-segmentation et de sous-segmentation, l'algorithme doit admettre des élisions et des insertions de phonèmes. Le résultat n'est plus un booléen indiquant la validation ou le rejet d'une hypothèse, mais un score de reconnaissance qui

constitue une mesure de certitude d'apparition du mot de référence. Ce score est calculé en appliquant une pénalité à la suite de chaque élision ou insertion. Dans le cas où le phonème recherché n'apparaît pas parmi les choix possibles, la solution à mettre en œuvre consiste à pénaliser la reconnaissance en fonction du degré de ressemblance entre le phonème recherché et les phonèmes proposés. Le résultat de l'analyse morphologique est un ensemble de mots ; à chaque mot sont associés un score de validité et un repère sur l'axe temporel de l'énoncé. Cette structure est représentée à l'aide d'un treillis de mots.

Organisation du lexique

Le système morphologique rigoureux a sûrement joué un rôle important dans l'organisation originale du lexique arabe. La conception du lexique peut se faire de différentes manières et sa réalisation est déterminée par l'architecture globale du système. Deux attitudes se dégagent selon l'objectif qu'on voudrait atteindre.

- L'une consiste à construire explicitement le lexique, celui-ci doit inclure toutes les formes fléchies des mots avec leurs informations relatives. La construction d'un tel dictionnaire peut être obtenue à l'aide de deux générateurs : un conjugueur de verbes et un dérivateur de noms. La génération se fait à partir d'une liste de racines auxquelles sont associées les informations relatives à la génération. Les changements concernant l'assimilation et la dissimilation seront résolus à l'aide de règles de réécriture [6]. Le problème se pose pour les applications faisant référence à un large vocabulaire ou dans le cas de la construction d'un lexique indépendant des applications. Un autre problème se pose au niveau de la construction du lexique, il s'agit de définir les informations à associer aux unités lexicales et les méthodes permettant de les exploiter par les niveaux morphologique, syntaxique, sémantique et phonétique.
- La seconde approche consiste à créer une partition sur l'ensemble des mots et à n'enregistrer dans les lexiques que les informations permettant de vérifier l'appartenance d'un mot à une classe d'équivalence. Chaque entrée du lexique constitue une classe d'équivalence à laquelle est associée une liste de valeurs grammaticales qui correspond à l'union des valeurs grammaticales des mots appartenant à cette même classe. Dans ce cas, l'analyse d'un mot se traduit par la recherche de préfixes, suffixes et schèmes vocaliques et puis la vérification de l'affinité entre les différents éléments. Cette idée doit, d'une part supposer que la formulation des règles émises par les grammairiens est facile à mettre en œuvre, d'autre part savoir utiliser les règles et pouvoir juger leurs limites. nous proposons d'obtenir ces lexiques par apprentissage [1], [10] .

Les attributs sémantiques : Le système de dérivation n'est pas seulement formel, mais il propage aussi des valeurs sémantiques.

Les objets qu'on cherche à représenter sont les entités linguistiques : mots, phrases, schèmes, racines, etc. et leur description relative. La caractérisation de ces objets nous permet de définir des subdivisions dans cet ensemble d'objets. Chaque subdivision définit un univers qui sera appelé par la suite concept. Les entités linguistiques seront appelées individus. Les concepts sont définis soit en extension, en énumérant ses individus avec leur description, c'est notamment le cas des mot-outils et des pronoms qui sortent du système de la racine, soit en intention à partir de règles d'inférence.

La transcription phonétique-orthographique : Le lexique est utilisé pour extraire la représentation phonétique d'un mot à partir de sa forme orthographique. Le passage d'une forme à une autre doit prendre en compte certaines caractéristiques de la langue en particulier :

- la quantité de la voyelle,
- la gémination,
- la nounation,
- l'assimilation,
- la prononciation du /ta/ /marbuta/ (liée) à la fin des noms et des adjectifs,
- les mots irréguliers, etc.

L'absence d'un clavier bilingue nous contraint à négliger cette phase et à travailler sur la structure phonétique interne du mot. Il arrive que le module syntaxico-sémantique demande la validation d'un ensemble de mots ayant la même fonction syntaxique ou possédant des traits sémantiques communs. Afin de permettre une sélection selon ces critères, il est indispensable d'affecter aux éléments du lexique des attributs syntaxiques et sémantiques.

4.2 Le module syntaxico-sémantique

La structure grammaticale de l'Arabe standard est rigide, et obéit à des règles qu'on peut formaliser. Cette rigidité n'est pas évidente lorsqu'il s'agit du langage parlé. Comme restriction, nous prenons, dans un premier temps, une grammaire artificielle qui engendre la plus grande partie des phrases. Le rôle du module syntaxico-sémantique est d'étudier les règles à suivre pour la construction des phrases, c'est à dire l'ordre dans lequel les mot doivent être disposés et les rapports grammaticaux qui doivent exister entre eux. La finalité du module est de produire une ou plusieurs structures syntaxico-sémantiques correspondant à un énoncé en langage pseudo-naturel. Ce modèle a la particularité de

proposer une représentation sémantique de chaque énoncé selon une structure composée d'un prédicat — dérivé d'une primitive — et d'un certain nombre d'arguments de cette primitive. Le traitement syntaxique s'appuie sur une base de connaissances sous forme de règles de production. Ce traitement construit une structure syntaxique intermédiaire. Le passage à une représentation conforme au modèle syntaxico-sémantique est assuré par l'application des règles de correspondance.

4.3 Le module prosodique

La prosodie est considérée comme étant la ponctuation du message parlé, le module prosodique détient un double rôle dans le système :

- il contribue, d'une part, à déterminer la signification de l'énoncé par l'émission des hypothèses concernant la forme de la phrase : cette information vient en complément des données destinées au processeur syntaxico-sémantique qui s'en sert pour lever certaines ambiguïtés.
- il s'efforce, d'autre part, de localiser des marqueurs prosodiques qui indiquent, sur le plan morphologique, une frontière de mot et, sur le plan syntaxique, une frontière de syntagme. Ainsi, les hypothèses morphologiques ou syntaxiques, qui entrent en contradiction avec les informations prosodiques, peuvent être abandonnées par le processeur correspondant. Par conséquent, le module prosodique doit retenir uniquement les marqueurs non-ambigus, c'est-à-dire ceux qui sont détectés avec un taux de certitude suffisamment élevé. Le travail revient à détecter la position de l'accent dans le mot et utiliser les règles qui régissent la présence de l'accent dans les mots arabes. Les paramètres acoustiques utilisés sont la durée, l'énergie et surtout l'évolution de la fréquence fondamentale ou pitch.

L'analyse de la courbe de variation de la fréquence fondamentale permet de classer les phrases de l'Arabe en cinq classes :

1. les phrases déclaratives : ce type de phrases est le plus fréquent et leur prononciation est un modèle pour le pitch.
2. les commandes : il s'agit d'un ordre, où la phrase est centrée autour du verbe conjugué à l'impératif. Le pitch est élevé au niveau du verbe même.
3. les questions : le pitch est plus élevé au niveau du mot de la question. Ce dernier se trouve généralement au début de la phrase.
4. les appels : ils sont constitués de l'article d'appel suivi de l'appelé. Les phrases sont généralement courtes et le pitch est pratiquement le même que pour les phrases déclaratives.

5. les exclamations : le pitch est plus élevé à la fin du mot principal de la phrase.

L'extraction du modèle du pitch permet de déterminer la nature de la phrase prononcée. Par ailleurs, dans une analyse prosodique faite sur des phrases affirmatives, il se dégage les effets intonatifs suivants :

- Toutes les phrases ont un schéma intonatif final descendant.
- Le schéma intonatif se réalise en fonction de la structure syntaxique de la phrase.
- Chaque mot de la phrase conserve son accent de l'état isolé.
- L'attaque intonative d'une phrase se situe sur la dynamique de base (120 Hz) pour les locuteurs masculins.

5 Conclusion

Nous avons présenté dans cet article une architecture d'un système de reconnaissance de l'Arabe moderne. Un système de reconnaissance de la parole continue dans un contexte multilocuteur. La partie décodage phonétique est en grande partie opérationnelle [8]. Il reste à introduire de nouvelles connaissances pour améliorer le pourcentage de reconnaissance phonétique, acquérir les connaissances nécessaires et développer les algorithmes de l'analyseur linguistique et assurer l'interaction entre les différents modules du système.

Références

- [1] A. Andreewsky. *Apprentissage, analyse automatique du langage, application à la documentation*. Document de linguistique quantitative, Paris, 1973.
- [2] S. H. Al. Ani. *Arabic Phonology. An Acoustical and Physiological Investigation*. Mouton & Co N.V., 1970.
- [3] A. Benhamouda. *Morphologie et syntaxe de la langue arabe*. Société Nationale d'édition et de Diffusion, 1983.
- [4] A. Betari. Artificial Intelligence Techniques for the Recognition of Standard Arabic. *Congrès international sur l'informatique*, pages 867–874, 1990.
- [5] R. Blachere and M Gaudefroy-Bemombynes. *Grammaire de l'Arabe classique*. Maisonneuve & larose, 1975.
- [6] F. Debili and L. Zouari. Analyse morphologique de l'Arabe écrit voyellé ou non fondée sur la construction automatisée d'un dictionnaire arabe. In *proceedings of Cognitiva'85*, pages 87–93, Paris, 1985.
- [7] M. Djoudi. Contribution à l'étude et à la reconnaissance automatique de la parole en arabe standard. Thèse de Doctorat de l'Université de Nancy 1, 1991.
- [8] M. Djoudi. Assessment of Acoustic Phonetic Decoder for Standard Arabic. In *13th National Computer Conference*, volume 2, pages 761–770, Riyadh Saudi Arabia, 28 November - 2 December, 1992.
- [9] M. Djoudi and J. P. Haton. The SAPHA Acoustic Phonetic Decoder System for Standard Arabic. In *1990 International Conference on Spoken Language Processing*, Kobe, Japan, 18-22 November, 1990.
- [10] C. Flhur. Algorithme à apprentissage et traitement automatique des langues. Thèse d'Etat, Université de Paris sud, 1977.
- [11] S. Ghazali. Elements of Arabic Phonetics. In *Applied Arabic Linguistics and Signal & Information Processing*, pages 51–58. Hemisphere publishing corporation, 1987.