

# Reconnaissance automatique de la parole en Arabe Moderne Standard

Mahieddine DJOUDI

L'Arabe moderne, dite aussi littéraire ou standard a fait l'objet de plusieurs travaux anciens ou récents, ayant trait soit à l'aspect phonétique soit à la composante linguistique de la langue [1] [2] [3] [4] [6] [8] [9] [10] [11] [12]. Toutefois, le problème de reconnaissance automatique n'a été que très peu abordé jusqu'à présent.

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 bien sûr des particularités de la langue. Au niveau phonétique, l'Arabe est caractérisée par la présence des consonnes emphatiques, pharyngales et glottales et l'existence d'une opposition brève-longue des voyelles. Au plan morphologique, la langue possède un système régulier basé sur la notion de racine qui consitue le pilier du dictionnaire arabe. Il devient clair que la conception d'un système de reconnaissance fait apparaître deux grands sous systèmes : un système de décodage acoustico-phonétique et un système de traitement des informations linguistiques. 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 linguistique 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 celle provenant de la définition du langage, détermine la signification du message oral [5].

## 1 Le décodeur acoustico-phonétique

C'est le premier niveau du processus de compréhension de la parole. Son rôle est de construire la description phonétique d'un énoncé, à partir du signal acoustique numérisé. Les principales étapes du système sont :

### 1.1 La segmentation du signal

Elle consiste à segmenter le signal de parole en grandes classes phonétiques [7], en utilisant des algorithmes non contextuels et reposant sur des critères simples. Le but essentiel de la segmentation est d'offrir un cadrage pour le module d'étiquetage et réduire ainsi l'explosion combinatoire lors de la reconnaissance.

Nous avons retenu quatre grandes classes :

- les voyelles { /a/, /i/, /u/, /aa/, /ii/ et /uu/ },
- les plosives { /t/, /k/, /ʔ/, /b/, /d/, /q/ et /ṭ/ },
- les fricatives { /z/, /f/, /θ/, /s/, /ʃ/, /χ/, /ħ/, /z/ et /ṣ/ }.
- le reste des phonèmes est mis dans la classe des "autres" qui comporte en particulier le vibrant /r/, le latéral /l/, les nasales /m/ et /n/, les semivoyelles /w/ et /y/ et les autres phonèmes de la langue.

### 1.2 Le calcul d'indices

L'extraction des indices phonétiques pertinents est une étape très importante dans le processus de décodage phonétique. Les indices actuellement utilisés sont : la durée du segment, le degré de voisement, la position et les caractéristiques de la barre d'explosion, le suivi de formant, les transitions formantiques, le centre de gravité et la limite inférieure du bruit de friction. Les indices sont calculés à partir de la représentation fréquentielle du signal.

### 1.3 L'étiquetage phonétique

Le module utilisé pour l'étiquetage phonétique est un système expert à base de règles de production. En partant des segments fournis par le module de segmentation, le module tente de trouver les bon phonèmes prononcés en utilisant les indices extraits lors de l'étape précédente 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.

## 2 Le décodeur linguistique

Le module de traitement des informations linguistiques reçoit en entrée un treillis phonétique et effectue la reconnaissance de la phrase traitée. Le résultat du système consiste en l'interprétation sémantique et l'orthographe de la phrase prononcée. Le décodeur linguistique comporte les modules suivants :

### 2.1 Le 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 appairer 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.

#### 2.1.1 Les 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.

##### Difficultés liées à la langue

- 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.
- Cas des mots homographes, ou une même chaîne recouvre deux notions différentes suivant le contexte, exemple /fi/ verbe impératif de /wafa/ et /fi/ préposition.

**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 décodeur phonétique 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 est 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.

### 2.1.2 Organisation du lexique

Le système morphologique rigoureux a sûrement joué un rôle important dans l'organisation originale du lexique arabe. La structure du lexique se traduit par son organisation en familles morphologiques ayant pour base étymologique la racine. Théoriquement, la dérivation se produit soit directement à partir de la racine, soit indirectement à partir d'une dérivée. Nous nous proposons de classer les dérivées en fonction de leur degré de relation avec la racine mère. Nous obtenons ainsi les dérivées primaires rattachées directement à la racine, les dérivées secondaires issues des primaires ; les tertiaires formées à partir des secondaires. Ainsi, le processus théorique de dérivation et le schéma morphologique qu'il engendre peuvent être représentés par une arborescence. Chaque racine donne naissance à un grand nombre de mots qui en dérivent par application des règles morphologiques.

**Les attributs sémantiques** 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.

**La transcription phonétique-orthographique** Le lexique est utilisé pour extraire la (les) représentation(s) phonétique(s) 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 (brève ou longue),
- la gémination et la nounation (/tanwin/),
- l'assimilation du /l/ devant les consonnes solaires,
- la prononciation du /ta/ /marbuta/ (liée) à la fin des noms et des adjectives,
- les mots irréguliers.

L'absence d'un clavier bilingue nous contraint de négliger cette phase et de travailler sur la représentation interne du mot.

## 2.2 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 de module syntaxico-sémantique est d'étudier les règles à suivre pour la construction des phrases, c'est à dire l'ordre dans lequel les mots 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. 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.

## 2.3 Le module prosodique

Ce module 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. Le travail revient à détecter la position de l'accent dans le mot et d'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.

En Arabe, il existe cinq types de phrases :

- les phrases déclaratives,
- les commandes,
- les questions,
- les appels,
- et les exclamations.

L'examen de la courbe de la fréquence fondamentale des phrases permet d'avoir une idée sur la nature de la phrase et de dégager les conclusions suivantes ;

- L'attaque intonative d'une phrase se situe sur la dynamique de base (120 Hz).
- Chaque mot de la phrase conserve son accent de l'état isolé.
- Il existe deux classes de schémas intonatifs qui se réalisent en fonction de la structure syntaxique de la phrase.
- Toutes les phrases ont un schéma intonatif final descendant.

### 3 Conclusion

Nous avons présenté dans cet article, une modeste contribution à la reconnaissance analytique de l'Arabe moderne. Les résultats actuels sont prometteuses. Pour trois locuteurs masculins, Le décodeur acoustico-phonétique arrive à faire la reconnaissance des phonèmes avec un taux global de l'ordre de 70%. Il reste à introduire de nouvelles connaissances pour améliorer le pourcentage de reconnaissance phonétique et développer les algorithmes du décodeur linguistique. Une fois opérationnel, le système peut être considéré comme le noyau d'une machine à dicter vocale, ou bien la partie la plus importante d'un système de dialogue oral homme-machine.

### Références

- [1] S. H. Al. Ani. *Arabic Phonology. An Acoustical and Physiological Investigation*. Mouton & Co N.V., 1970.
- [2] A. Benhamouda. *Morphologie et syntaxe de la langue arabe*. Société Nationale d'édition et de Diffusion, 1983.
- [3] R. Blachere and M Gaudefroy-Bemombynes. *Grammaire de l'Arabe classique*. Maisonneuve & Larose, 1975.
- [4] J. Cantineau. *Cours de phonétique arabe*. Librairie Klincksieck, 1960.
- [5] M. Djoudi, D. Fohr, and J. P. Haton. MARS : Un système de reconnaissance de l'Arabe moderne. In *Actes des 18<sup>ème</sup> Journées d'Etudes sur la Parole*, pages 217–221, Montréal, Mai, 1990.
- [6] M. Djoudi, D. Fohr, and J. P. Haton. Phonetic Study for Automatic Recognition of Arabic. In *Proceedings of European Conference on Speech and Technology*, volume 2, pages 268–271, Paris, September 1989.
- [7] 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.
- [8] H. Fleisch. *Traité de philologie arabe : Préliminaires, phonétique, morphologie nominale*, volume 1. Imprimerie Catholique Beyrouth, Beyrouth, 1961.
- [9] Ibn Jinni. *Sirr Sinaa'at Al I'raab*. Mustapha Al Halabi, 1954.
- [10] A. Roman. *Etude de la phonologie et de la morphologie de la Koine arabe*, volume 1. Université de Provence, Aix en Provence, 1983.
- [11] S. De Sacy. *Grammaire arabe*. De Sacy, 1810.
- [12] Sibawayh. *EL KITAB, traité de grammaire arabe*. H. Derembourg, 1889.