

Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF - ISO 24613

Aïda KHEMAKHEM¹, Bilel GARGOURI¹,
Abdelhamid ABDELWAHED², Gil FRANCOPOULO³

¹Laboratoire MIRACL, FSEG-SFAX B.P. 1088, 3018 SFAX – TUNISIE

²Unité de recherche LSCA, FLSH-SFAX B.P. 553, 3018 SFAX – TUNISIE,
³INRIA-Loria

khemakhem.aida@gnet.tn, Bilel.Gargouri@fsegs.rnu.tn
abdelhamid.abdelwahed@yahoo.fr, gil.francopoulo@wanadoo.fr

Résumé. Dans cet article, nous spécifions les paradigmes de flexion des verbes arabes en respectant la version 9 de LMF (Lexical Markup Framework), future norme ISO 24613 qui traite de la standardisation des bases lexicales. La spécification de ces paradigmes se fonde sur une combinaison des racines et des schèmes. En particulier, nous mettons en relief les terminaisons de racines sensibles aux ajouts de suffixes et ce, afin de couvrir les situations non considérées dans les travaux existants. L'élaboration des paradigmes de flexion verbale que nous proposons est une description en intension d'*ArabicLDB* (Arabic Lexical DataBase) qui est une base lexicale normalisée pour la langue arabe. Nos travaux sont illustrés par la réalisation d'un conjugueur des verbes arabes à partir d'*ArabicLDB*.

Abstract. In this paper, we specify the inflected paradigms of Arabic verbs with respect to the version 9 of LMF (Lexical Markup Framework) which is the expected ISO 24613 standard dealing with the standardisation of lexical databases. The specification of these paradigms is based on a combination of schemes and roots. In particular, we highlight the role of root endings that is not considered in other researches and that may generate erroneous forms while concatenating suffixes. The development of verbal inflected paradigms that we propose is an intentional component of *ArabicLDB* (Arabic Lexical DataBase) which is a normalized Arabic lexical database that we developed according to LMF. Our works are illustrated by the realization of a conjugation tool for Arabic verbs using *ArabicLDB*.

Mots-clés : langue arabe, paradigmes de flexion verbale, base lexicale, norme ISO 24613, LMF, lexical markup framework, conjugueur des verbes arabes.

Keywords: Arabic, inflected paradigms of verb, lexical database, norm ISO 24613, LMF, lexical markup framework, conjugation of arabic verbs.

1 Introduction

L'arabe est une langue à la fois dérivationnelle et flexionnelle (Blachère et al, 1975). Elle se caractérise par une morphologie assez complexe à la manière de la conjugaison de ses verbes

qui génère plusieurs formes fléchies variant d'un verbe à un autre. Ainsi, s'accroît le besoin de modéliser les paradigmes de flexion de l'arabe afin de limiter l'espace de représentation des formes fléchies lors de la construction d'une base lexicale.

Dans le présent papier, nous nous intéressons à la modélisation des paradigmes de flexion des verbes arabes en vue de construire une description en intension des verbes dans une base lexicale. Nos investigations sont fondées sur les principaux travaux existants dans ce domaine (El-Dahdah, 1999), (Ammar et al, 1999), (Abdelwaheh, 1996) et (Ibn elqataa, 2003).

Ces travaux classent les verbes selon des critères qui se fondent sur les consonnes sensibles à la modification des voyelles voisines. Néanmoins, nous avons pris en considération certains cas liés à l'ajout d'un suffixe qui peuvent engendrer des formes erronées. Partant de cette constatation, nous proposons d'adapter la classification des racines selon les besoins du traitement automatique en vue de couvrir des cas omis dans la quasi-totalité des travaux sur le sujet.

En ce qui concerne la modélisation et la réalisation, nous profitons du travail en cours au sein de l'ISO de la spécification LMF (Lexical Markup Framework) (Francopoulo, 2003) qui propose une représentation lexicale standard pour les langues les plus utilisées tout en couvrant, entre autres, la représentation des paradigmes de flexion. Ce projet est en cours de validation par le comité TC37/SC4 de l'ISO sous la référence ISO-24613 (Francopoulo, 2006). Comme illustration de l'application de LMF sur le cas de la langue arabe, nous présentons la base *ArabicLDB* qui se limite dans sa version actuelle à la représentation du niveau morphologique (Khemakhem et al, 2006), (Khemakhem, 2006).

Le présent travail est réalisé selon les recommandations de la version 9 de LMF appliquées à la base *ArabicLDB* représentant une description des paradigmes de flexion verbale.

Dans la première partie, nous présentons les travaux actuels de modélisation des paradigmes de flexion des verbes arabes. Ensuite, nous donnerons une idée sur le projet de normalisation LMF, en s'intéressant en particulier à l'extension des paradigmes de flexion. Dans la section suivante, nous présenterons la modélisation des paradigmes verbaux de la langue arabe selon LMF. Puis nous spécifierons les verbes types de la langue arabe. Enfin, nous terminerons par l'implémentation des paradigmes de la base *ArabicLDB* avec un descriptif du conjugeur réalisé.

2 Aperçu sur les travaux de modélisation des paradigmes des verbes arabes

2.1 Caractéristiques des verbes arabes

Dans la langue arabe, il y a plus de 16 000 verbes. Un verbe peut avoir 109 formes fléchies quand il admet à la fois la voix active et passive et 57 formes fléchies (FF) lorsqu'il n'admet que la voix active (Ammar et al, 1999). En général, les verbes sont des formes dérivées à partir d'une racine et d'un schème. La racine est purement consonantique qui peut être formée soit par trois consonnes formant des verbes trilitères, soit par quatre consonnes formant des verbes quadrilitères. Le schème peut être considéré comme une représentation formelle constituée par 3 ou 4 consonnes (ل, ع, ف) qui sont totalement vocalisées, ou comme un moule sur laquelle coule la racine (Baloul, 2003). En totalité, il y a 19 schèmes verbaux qui peuvent être soit nus, soit augmentés dérivant de trois consonnes de la racine par modification des voyelles, par redoublement de la deuxième lettre de la racine, par adjonction et même par

intercalation d'affixes (préfixe, infixe, suffixe). Les verbes augmentés se conjuguent avec les mêmes préfixes et suffixes que le verbe sans augment. De ce fait, une racine peut générer au maximum 19 verbes et les schèmes correspondants peuvent donner 22 modèles de conjugaison différents. En effet, il y a le schème فَعَلَ [fa'ala] qui peut avoir trois alternances différentes de conjugaison selon la nature de la voyelle de la 2^e consonne de la racine de l'inaccompli يَقَعُلُ [yaf'ulu], يَقَعُلُ [yaf'īlu], et لَيَقَعُلُ [yaf'alu]. En plus, le schème فَعِلَ [fa'ila] peut donner deux alternances différentes de conjugaison pour la même raison (El-Dahdah, 1999).

La conjugaison d'un verbe arabe consiste à engendrer l'ensemble des formes que peut prendre ce verbe à la voix active ou passive. Ces formes varient selon le mode et l'aspect (accompli, inaccompli indicatif, inaccompli subjonctif, inaccompli apocopé et l'impératif), comme elles varient selon le nombre (singulier, duel, pluriel) et le genre (masculin, féminin). Elles varient aussi selon la personne ou les personnes représentées par le sujet : كَاتَبْتُ [katabtu], كَاتَبْتُمْ [katabtumâ]...

Généralement, les règles de conjugaison d'un verbe arabe peuvent engendrer les modifications suivantes : le changement des voyelles, l'ajout d'un préfixe ou d'un suffixe, et la suppression d'un *hamza wasliya* au début ou d'une voyelle à la fin. De plus, ces modifications peuvent avoir recours à des règles phonologiques qui opèrent sur certaines formes verbales particulières.

2.2 Les principales propositions du domaine

Les travaux sur la langue arabe sont très nombreux mais très dispersés comparés à ceux d'autres langues comme l'anglais ou le français. Parmi les principaux travaux sur les verbes arabes, nous citons « معجم تصريف الأفعال العربية » (El-Dahdah, 1999), « Les verbes arabes » (Ammar et al, 1999), « قراءة في التصريف العربي. بنية الفعل » (Abdelwaheh, 1996) et « باتك » (Ibn elqataa, 2003) qui traitent un nombre important de verbes, mais qui sont destinés à une utilisation manuelle.

Signalons que les travaux sur les verbes possèdent quelques points communs, comme la distribution des verbes et la détermination de leur nature, ainsi que l'idée principale qui consiste à traiter les règles phonologiques comme des règles de conjugaison ou morphologique. Selon ces auteurs, les causes de déclenchement des règles phonologiques sont limitées par la nature des consonnes de la racine. Ce qui a pour effet de déterminer un classement des paradigmes de flexion et des verbes. Selon cette optique, deux principales classes de verbe sont distinguées: les verbes sains (حيصص) qui ne comportent pas de lettres défectueuses, et les verbes défectueux (معتل) qui contiennent une ou deux lettres défectueuses causant des altérations importantes au cours de la conjugaison. Un verbe sain peut contenir la lettre *hamza* ou *šadda* qui peut engendrer des conjugaisons irrégulières.

Bien qu'il y ait des points communs, les différences impliquent des conséquences importantes pour la représentation et la couverture : Ammar traite seulement 10 000 verbes, alors que El-Dahdah traite plus de 16 000 des verbes arabes. Notons que ce dernier auteur couvre la plupart des verbes quadrilittères ainsi que les verbes trilitères rarement utilisés.

La complexité morphologique de la langue arabe est constatée entre autres dans l'utilisation des règles phonologiques au cours du calcul d'une forme fléchie. Par conséquent, les linguistes ont proposé la classification des verbes selon la racine et ils ont valorisé les consonnes (la lettre *hamza*, les lettres défectueuses ou le signe *šadda*) sensibles aux changements de leurs voyelles (précédente et suivante). Cependant, l'automatisation directe

de cette classification, peut nous donner des formes erronées. Pour éviter l'apparition de ces formes, il faut valoriser les consonnes sensibles à l'ajout d'un suffixe (les terminaisons qui peuvent être fusionnées avec la première consonne du suffixe) (Khemakhem, 2006).

Lemme	حَزَنَ	[ħazana]
Forme Fléchie intermédiaire avec هُنَّ	حَزَنْتِنَ	
Forme Fléchie finale avec هُنَّ	حَزَنْتِنَ	[ħazanna]

Figure 1 : Exemple de fusion d'une terminaison et suffixe utilisant une règle phonologique

Dans cet exemple, si ce verbe est traité comme les autres verbes sains, la forme intermédiaire « حَزَنْتِنَ » [ħazanna] sera une forme erronée qui nécessitera l'intervention d'une règle phonologique pour donner la forme finale « حَزَنْتِنَ » [ħazanna].

3 LMF : future norme ISO 24613

3.1 Présentation générale

LMF (Lexical Markup Framework) est un projet en cours de validation par le sous comité TC-37/SC-4 de l'ISO sous la référence ISO-24613. Le projet traite de la normalisation des bases lexicales à large couverture (Francopoulo et al, 2006). Ce projet est né du besoin exprimé par les différentes délégations nationales de pouvoir représenter, échanger et fusionner les bases lexicales actuelles et futures. Notons que le besoin de fusion est celui qui est le plus souvent exprimé. La totalité de LMF est spécifiée à l'aide d'UML (Unified Modeling Language) qui s'impose maintenant comme le standard de modélisation des structures. Le format d'échange est défini en XML.

Un lexique LMF doit comporter une partie noyau et un ou plusieurs modules d'extension qui sont spécifiés par des "packages" UML. Le noyau traite des notions de lexique (en tant que conteneur), de mot, de forme et de sens. Le noyau est à considérer à la fois comme le squelette du lexique et comme le dénominateur commun à plusieurs lexiques, ce qui est essentiel pour les opérations de fusion. Le noyau est obligatoire. Autour de ce dernier, le créateur du lexique peut définir des modules d'extension pour la morphologie, les paradigmes de flexion, la syntaxe, la sémantique et/ou les notations multilingues.

LMF définit la structure du lexique mais n'en précise pas les détails. La structure est décorée par des couples attribut-valeur à prendre dans un registre de catégories de données (RCD) qui est géré par l'ISO dans le cadre de l'ISO-12620 (<http://syntax.inist.fr/>) (Romary et al, 2003).

Ainsi, LMF permet deux axes de variation : d'une part le choix des modules d'extension et d'autre part le choix des catégories de données. Par exemple, un gestionnaire de lexique de morphologie n'aura pas à s'embarrasser des mécanismes descriptifs de la syntaxe, s'il n'en éprouve pas le besoin. De même, ce même gestionnaire sélectionnera les valeurs /masculine/ et /feminine/ pour le français alors qu'il choisira /masculine/, /feminine/ et /neuter/ pour l'allemand.

3.2 Le module d'extension des paradigmes de flexion

Le module des paradigmes proposé dans LMF est une description simple des modifications qui sont apportées au lemme pour calculer toutes les formes fléchies d'une entrée lexicale.

Cette description utilise huit opérations (/add/, /addBefore/, /addAfter/, /remove/, /removeAfter/, /removeBefore/, /substitute/, /copy/) qui sont liées à un calculateur de forme fléchie spécifique à une combinaison de traits morphologiques (Francopoulo et al, 2006).

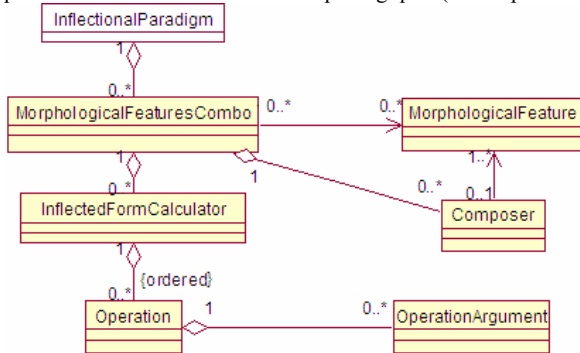


Figure 2 : L’extension des paradigmes de flexion selon LMF

Les paradigmes de flexion comportent plusieurs instances de la classe *Morphological FeaturesCombinator*. Ces dernières sont relatives à une classe abstraite qui fait le lien entre des traits morphologiques et un ou plusieurs *InflectedFormCalculators* qui regroupent des *Operations* et des *OperationArguments* permettant de calculer les FF correspondantes.

4 La base lexicale ArabicLDB

ArabicLDB (Khemakhem, 2006) est une nouvelle base lexicale arabe, qui est conforme à la révision 9 de la future norme LMF. Elle est implémentée en XML en respectant la DTD proposée dans cette révision (Francopoulo et al, 2006). Le choix a été pris de décrire explicitement toutes les FF des noms et de décrire les verbes en intension à l’aide des paradigmes de flexion.

Les entrées lexicales d’*ArabicLDB* peuvent être des formes dérivées (nominales ou verbales), des noms non dérivés et des particules, qui sont totalement vocalisées. Pour chaque partie du discours, un ensemble de catégories de données est sélectionné à partir du RCD quand c’était possible. Certaines valeurs sont absentes du registre actuel. Nous les avons utilisées et testées, puis, nous avons demandé leur ajout dans le registre de l’ISO.

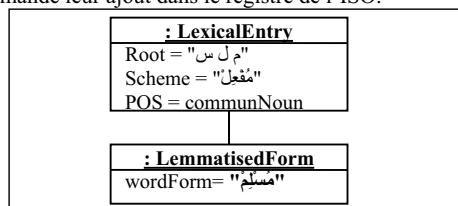


Figure 3 : Exemple d’une entrée lexicale

Les entrées lexicales portent un identifiant et l’information de partie du discours. Ce choix implique la création de deux entrées pour des formes identiques à parties de discours distinctes (كَمْ : pronom interrogatif vs. كَمْ : pronom allusif (كَتَابِيَة)). En contrepartie, des

homonymes à partie de discours identique (ساعة : une heure vs. ساعة : une montre) ne sont pas traités dans deux entrées différentes.

Pour la langue arabe, le verbe et la plupart des noms sont identifiés par une racine et un schème. En effet, nous utilisons /root/ et /scheme/ comme des couples attribut-valeur de la classe *LexicalEntry*, qui est la classe permettant de représenter la notion de mot.

Actuellement, nous avons un outil d'alimentation de cette base et d'autres outils pour l'exploitation. Nous avons des modules d'acquisition des paradigmes (256) et des entrées lexicales (plus de 16 000 verbe, 500 particules...). Pour l'exploitation, nous avons des modules de recherche et de conjugaison des verbes.

5 Les paradigmes de flexion verbale de l'arabe selon LMF

Les paradigmes de flexion sont des représentations communes à un grand nombre de mots dans une optique de description en intention. Ils factorisent la connaissance linguistique sous forme d'un prototype de conjugaison. Ils facilitent donc la maintenance et réduisent la taille de la base de données. Cependant, le chevauchement entre les différents niveaux linguistiques dans la langue arabe, notamment morphologique et phonologique, et l'absence de la phonologie dans la norme, montrent que ces paradigmes nécessitent une investigation très détaillée. En effet, nous suivons une démarche qui procède en trois étapes: la classification, l'identification des verbes types et l'application des opérations (selon LMF).

5.1 Classification des verbes

L'objectif de cette phase était l'étude des consonnes qui déterminent une règle phonologique. Après l'étude des classifications des racines déjà proposées dans la littérature, nous les avons précisé en valorisant les terminaisons (les lettres ت et ن) qui peuvent se fusionner avec la première consonne du suffixe. La Figure 4 donnée ci-après, illustre la considération des terminaisons lors de la classification des verbes. Dans cette figure, la sous classe des verbes sains (سالم) sera constituée de trois sous classes : la première (*) regroupe les verbes qui ne terminent ni par la lettre ت ni par la lettre ن, la deuxième (#ن) regroupe les verbes ayant la terminaison ن et la troisième (#ت) regroupe les verbes ayant la terminaison ت.

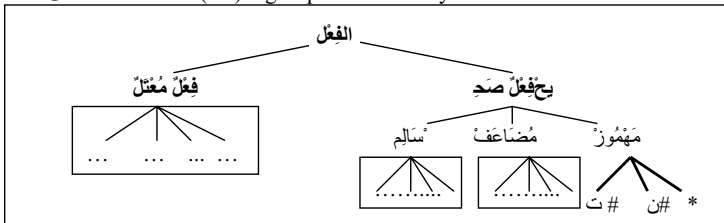


Figure 4 : Rôle des terminaisons dans la classification des verbes

En résultat, nous avons dégagé 42 classes de racines verbales en considérant les terminaisons qui peuvent se fusionner avec la première consonne du suffixe.

5.2 Identification des verbes types

Le modèle LMF n'a pas d'extension phonologique, ce qui nous amène à traiter les transformations phonologiques comme des flexions morphologiques. Et nous n'avons pas que 22 modèles de conjugaison puisqu'il faut considérer les différents types de racine possibles.

Par ailleurs, nous avons combiné les modèles de conjugaison (les 22 classes de schème) et les classes de racine. Ensuite, nous avons optimisé les résultats pour éviter la redondance des paradigmes et l'apparition de paradigmes superflus: il y a des schèmes qui revêtent un sens assez particulier et qui sont utilisés uniquement avec quelques classes de racine.

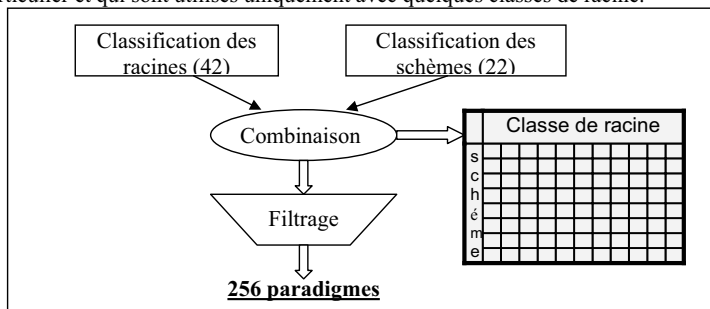


Figure 5 : Démarche de préparation des paradigmes verbaux

Après la combinaison et l'optimisation des combinaisons des classes de racine et des classes de schème, nous avons trouvé 256 paradigmes qui correspondent aux verbes types (Khemakhem, 2006).

5.3 Application des opérations proposées par LMF pour le cas de l'arabe

Dans cette section, il s'agit d'étudier la couverture de toutes les transformations possibles au cours de la conjugaison des verbes par les opérations proposées par LMF.

Nous rappelons que les règles de conjugaison d'un verbe arabe peuvent engendrer les modifications suivantes : le changement des voyelles, l'ajout d'un préfixe ou d'un suffixe, et la suppression d'un hamza wasliya au début ou une voyelle à la fin. Nous avons utilisé les opérations proposées dans LMF pour appliquer les règles de conjugaison de l'arabe. Comme illustration, nous présentons dans la Figure 6, l'exemple de calcul de la forme fléchie « **يَسْتَقْبِلُ** » [yastaqbilu] à partir du lemme « **اِسْتَقْبَلَ** » [istaqbala].

Nous avons encore des cas particuliers qui nécessitent des modifications particulières liées à l'application des règles phonologiques. Ces modifications sont :

- la transformation de la graphie de *hamza*, ce qui nécessite parfois les opérations de remplacement **substitute** ou de suppression **remove**.
- la transformation des lettres défectueuses, ce qui nécessite parfois les opérations de remplacement **substitute** ou de suppression **remove**.
- la transformation de *šadda* qui peut être supprimé par **remove**, la consonne impliquée dupliquée par **copy**. De plus, des voyelles peuvent être ajoutées par **add**.
- la fusion du suffixe avec la dernière consonne de la racine **ن** ou **ت**. Elle est assurée par la suppression **remove** d'une consonne avec sa voyelle et l'ajout **add** de *šadda*.

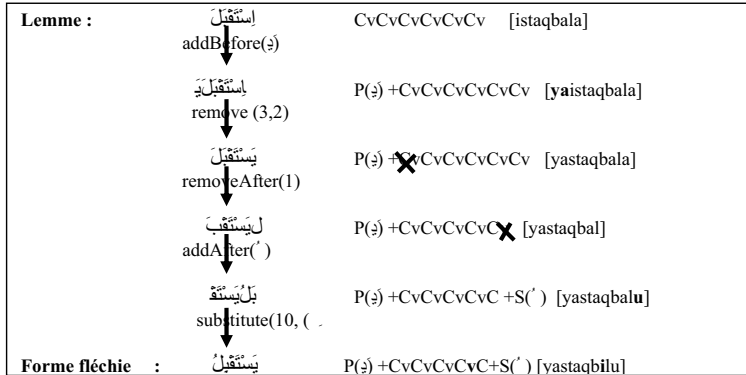


Figure 6 : Exemple d'un calcul d'une forme fléchie

D'après l'étude de tous les cas particuliers liés à la phonologie, nous pouvons conclure que les opérations proposées par la norme LMF couvrent toutes les flexions de la langue arabe.

6 Elaboration de la base des paradigmes

6.1 La base des paradigmes

Cette base comporte la description en XML des combinaisons des traits morphologiques pour chaque verbe type. Une interface appropriée a été mise en place pour faciliter la création d'un paradigme de flexion. L'utilisateur peut choisir une combinaison et spécifier les opérations en précisant les arguments associés. Comme illustration, nous présentons, dans la Figure 7, l'exemple de « اِسْتَقْبَلْ » [yastaqbilu] de la figure précédente.

```

<Lexicon> <InflectionalParadigm id="asIstaqbala">
<MorphologicalFeaturesCombo><MorphologicalFeature att="verbFormAspect" val="unaccomplished"/>
<MorphologicalFeature att="verbFormMood" val="indicative" />
<MorphologicalFeature att="voice" val="activeVoice" />
<MorphologicalFeature att="person" val="thirdPerson" />
<MorphologicalFeature att="grammaticalNumber" val="singular" />
<MorphologicalFeature att="grammaticalGender" val="masculine" />
<InflectedFormCalculator> <DC att="stem" val="0" />
<Operation> <DC att="graphicalOperator" val="addBefore" />
<OperationArgument> <DC att="chain" val="ي" /> </OperationArgument> </Operation>
<Operation> <DC att="graphicalOperator" val="remove" />
<OperationArgument> <DC att="pos" val="3" />
<OperationArgument> <DC att="number" val="2" /> </OperationArgument> </Operation>
<Operation> <DC att="graphicalOperator" val="removeAfter" />
<OperationArgument> <DC att="number" val="1" /> </OperationArgument> </Operation>
<Operation> <DC att="graphicalOperator" val="addAfter" />
<OperationArgument> <DC att="chain" val="" /> </OperationArgument> </Operation>
<Operation> <DC att="graphicalOperator" val="substitute" />
<OperationArgument> <DC att="chain" val="" /> </OperationArgument>
<OperationArgument> <DC att="pos" val="10" /> </OperationArgument> </Operation>
</InflectedFormCalculator></MorphologicalFeaturesCombo>...
</InflectionalParadigm>.....</Lexicon>
    
```

Figure 7 : Combinaison de traits morphologiques du paradigme du verbe "اِسْتَقْبَلْ"

6.2 Le module d'extension morphologique de la base ArabicLDB

Après l'élaboration des paradigmes qui se caractérisent par un identifiant, nous pouvons spécifier pour chaque verbe son paradigme de flexion au niveau du lemme. En plus, nous ajoutons une catégorie de données /havePassive/ qui est spécifique pour les verbes. Elle prend **no** si le verbe admet seulement la voix active et **yes** si le verbe admet les deux voix.

```

<lexicalEntry> <DC att="root" val="ب ت ك" />
<DC att="scheme" val="فعل" />
<DC att="pos" val="verb" />
<lemmatisedForm paradigm="asKataba"> <DC att="wordForm" val="كُتِبَ" />
<DC att="havePassive" val="yes" />
</lemmatisedForm>
</lexicalEntry>
    
```

Figure 8 : Exemple d'une entrée lexicale verbale d'ArabicLDB

Dans cet exemple, nous avons spécifié le paradigme du verbe "كُتِبَ" [kataba] comme attribut de ce lemme. Cet attribut joue le rôle d'une référence pour trouver le paradigme convenable à un verbe dans le module des modes de flexion. Ce paradigme peut être appliqué totalement sur ce verbe puisque la catégorie de données /havePassive/ prend la valeur **yes**. Autrement dit, ce verbe admet les deux voix (active et passive).

7 Mise en œuvre d'un conjugeur

Le conjugeur de la base *ArabicLDB* utilise le module des paradigmes de flexions pour engendrer toutes les formes fléchies d'un verbe donné à partir du lemme. Dans un premier temps, le programme accède à l'extension morphologique pour chercher l'identifiant du paradigme du verbe. Dans un deuxième temps, il accède à l'extension des modes de flexion pour importer le paradigme en question. Nous rappelons que ce paradigme est composé de plusieurs combinaisons de traits morphologiques dont chacune d'elles possède des opérations et des arguments (position, chaîne, etc.) qui permettent le calcul des formes fléchies.

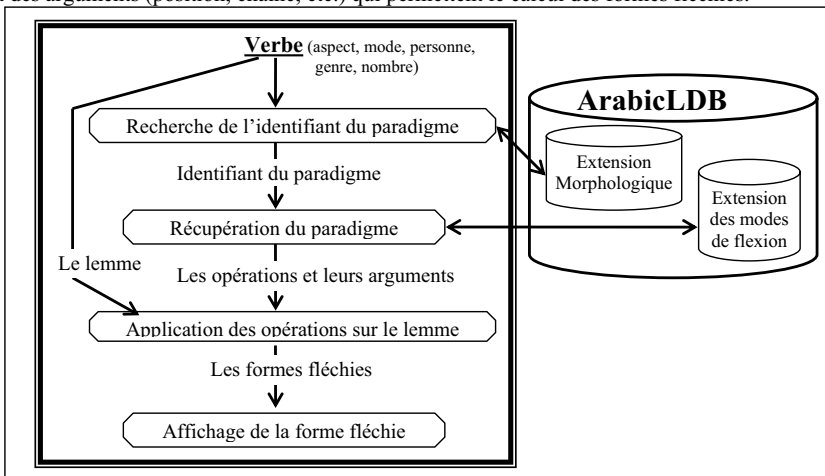


Figure 9 : Architecture du conjugeur

Notre conjugueur est composé de quatre modules :

- **Recherche de l'identifiant du paradigme** : c'est la première étape de ce conjugueur qui consiste à trouver l'identifiant du paradigme du verbe en question, en accédant à l'extension morphologique de la base *ArabicLDB*.
- **Récupération du paradigme** : cette étape utilise l'identifiant du paradigme pour accéder à l'extension des modes de flexion et importer les données de ce paradigme.
- **Application des opérations sur le lemme** : ce module permet la génération des formes fléchies en se basant sur les opérations et leurs arguments.
- **Affichage de la forme fléchie** : nous affichons la forme fléchie générée avec ses traits morphologiques.

8 Conclusion

L'absence d'une représentation intensionnelle robuste pour les verbes arabes, nous a poussé à étudier ce domaine. En plus, l'apparition d'une norme ISO (LMF) nous a encouragé à élaborer les paradigmes de flexion des verbes arabes selon cette norme sachant que nous avons déjà utilisé LMF pour élaborer la base morphologique *ArabicLDB*.

En ce qui concerne l'identification des verbes types, nous avons adopté les classifications proposées dans les travaux existants afin de combiner les classes résultantes avec les classes de schème tout en mettant en relief les terminaisons des racines sensibles à l'ajout des suffixes. Ensuite, nous avons filtré ces combinaisons en utilisant des critères d'ordre sémantique pour aboutir à la spécification de 256 paradigmes verbaux. Ces paradigmes forment la représentation en intension des verbes dans la base *ArabicLDB* qui comporte 16 000 verbes.

Un conjugueur a été mis en place comme outil d'exploitation et de vérification en s'appuyant sur *ArabicLDB*. Ce conjugueur est utilisé dans un contexte d'enseignement de l'arabe.

Références

- ABDELWAHED A. (1996).. كلية الآداب و العلوم الإنسانية بصفاس، تونس.
- AMMAR S., DICHY J. (1999). *Les verbes arabes* (Collection Bescherelle). Editions HATIER, Paris.
- BALOUL S. (2003). Développement d'un système automatique de la parole à partir du texte arabe standard voyellé. Thèse de doctorat de l'université du MAINE, Le Mans, France.
- BLACHERÉ R., GAUDEFROY-DEMOMBYNES M. (1975). *Grammaire de l'arabe classique*. Lieu : Edition Maisonneuve-Larose, Paris.
- EL-DAHDEH A. (1999). انجيل، شوريب، نورشان نانبل تبتكلم.
- FRANCOPOULO G. (2003). Proposition de normalisation de norme des lexiques pour le traitement automatique du langage. INRIA/LORIA-ACTION SYNTAXE, Version-1.3.
- FRANCOPOULO G., GEORGE M. (2006). ISO/TC 37/SC4 N130 Rev.9. Language resource management – Lexical markup framework (LMF).
- IBN ELQATAA (2003). شوريب، فيمعلعلا بتكفل ا راد .
- KHEMAKHEM A., GARGOURI B., ABDELWAHED A. (2006). LMF est-il convenable pour la langue arabe ? Actes de *Journées sur le Traitement Automatique de la Langue Arabe JTALA'06*, Rabat, Maroc.
- KHEMAKHEM A. (2006). "ArabicLDB : une base lexicale normalisée pour la langue arabe". Mémoire de master en Systèmes d'Information et Nouvelles Technologies. Lieu : FSEG, Sfax, Tunis. (<http://www.tagmatica.fr/doc/MemoireAida.pdf>).
- ROMARY L., WRIGHT S., FARRA S., GILLAM L. (2003). ISO TC 37/SC4 N055, Language resource management - Implementing a data category registry within ISO TC37.