

Modélisation normalisée LMF des dictionnaires électroniques éditoriaux de l'arabe

Feten BACCAR (1), Aïda KHEMAKHEM (1), Bilel GARGOURI (1),
Kais HADDAR (2), Abdelmajid BEN HAMADOU (3)

Laboratoire MIRACL

(1)FSEGS, B.P. 1088, 3018 Sfax, Tunisie

baccarf@yahoo.fr aida_khemakhem@yahoo.fr bilel.gargouri@fsegs.rnu.tn

(2)FSS, B.P. 802, 3038 Sfax, Tunisie

kais.haddar@fss.rnu.tn

(3)ISIMS, B.P. 242, 3021 Sakiet-Ezzit Sfax, Tunisie

abdelmajid.benhamadou@isimsf.rnu.tn

Résumé

Le présent papier s'intéresse à l'élaboration des dictionnaires électroniques arabes à usage éditorial. Il propose un modèle unifié et normalisé de ces dictionnaires en se référant à la future norme LMF (Lexical Markup Framework) ISO 24613. Ce modèle permet de construire des dictionnaires extensibles, sur lesquels on peut réaliser, grâce à une structuration fine et standard, des fonctions de consultation génériques adaptées aux besoins des utilisateurs. La mise en œuvre du modèle proposé est testée sur des dictionnaires existants de la langue arabe en utilisant, pour la consultation, le système ADIQTO (Arabic DIctionary Query TOols) que nous avons développé pour l'interrogation générique des dictionnaires normalisés de l'arabe.

Abstract

This paper is interested in the development of the Arabic electronic dictionaries of human use. It proposes a unified and standardized model for these dictionaries according to the future standard LMF (Lexical Markup Framework) ISO 24613. Thanks to its subtle and standardized structure, this model allows the development of extendable dictionaries on which generic interrogation functions adapted to the user's needs can be implemented. This model has already been carried out on some existing Arabic dictionaries using the ADIQTO (Arabic DIctionary Query Tool) system, which we developed for the generic interrogation of standardized dictionaries of Arabic.

Mots-clés : Dictionnaire électronique Arabe, usage éditorial, modèle normalisé, LMF, interrogation générique

Keywords: Arabic electronic dictionary, human use, standardized model, LMF, generic interrogation

1 Introduction

L'avènement de l'outil informatique et la prolifération des moyens de communication ont changé nos modes de stockage et d'accès aux informations. De ce fait, les dictionnaires électroniques à usage humain (ou éditorial) se trouvent libérés des contraintes de leurs versions papiers et répondent à un certain nombre de besoins pratiques d'actualisation et d'exploitation pour une population hétérogène d'utilisateurs (i.e., les apprenants, les experts journalistes ou écrivains et les chercheurs en linguistique) (Dendien & Pierrel, 2003), (Wooldridge, 1998), et (Arregi & al, 2002). Sur cette lancée, plusieurs dictionnaires électroniques à usage humain sont aujourd'hui disponibles sur le Web et/ou commercialisés sur des CD-ROM, notamment pour les langues occidentales.

Les travaux d'informatisation des dictionnaires occidentaux ont débuté depuis quelques décennies. L'objectif était de faire bénéficier les utilisateurs d'une aide éminemment précieuse, dans la mesure où ces dictionnaires électroniques permettent un accès sélectif et rapide aux informations. Ces services consistent, notamment, à retrouver un mot par son orthographe approximative, extraire une liste de termes spécialisés se rapportant à un domaine, rechercher une liste de mots sur des critères autres qu'alphabétiques, etc. Parmi les dictionnaires électroniques occidentaux, nous citons ici quelques exemples pour les langues française et anglaise : le Trésor de la Langue Française informatisé (TLFi) [<http://www.atilf.fr/tlfi>], le Dictionnaire de l'Académie Française [<http://www.atilf.fr/academie9>], le Nouveau Petit Robert électronique, le Bibliorom Larousse, Oxford English Dictionary [<http://www.oed.com>] et Collins Electronic English Dictionary & Thesaurus.

Par ailleurs, la diversité structurale des ressources dictionnaires disponibles a rendu l'échange et la fusion de leurs données incontestablement difficiles et complexes. De toute évidence, une représentation lexicale unifiée et normalisée est nécessaire et préalable à toute exploitation de ces ressources hors de leur propre contexte de conception (Francopoulo, 2004). Ainsi, deux principales propositions de standardisation de dictionnaires électroniques ont vu le jour, à savoir la TEI (Text Encoding Initiative) (Sperberg-McQueen & Burnard, 2005) et LMF (Lexical Markup Framework) (Francopoulo & George, 2008).

De ces travaux et réalisations, la langue arabe n'a pas pu profiter pleinement. L'élaboration de dictionnaires électroniques arabes, notamment à usage humain, est dans une phase qu'on peut qualifier de primaire. En effet, il appert que la quasi-totalité des dictionnaires électroniques arabes disponibles ne sont rien d'autre qu'une numérisation de fonds dictionnaires (sous format : ".doc", ".pdf" ou ".html"), non structurées et peu bénéfiques (Ait Taleb, 2005). Par ailleurs, les outils de consultation accompagnant ces dictionnaires offrent un seul type de recherche qualifié de primitif (recherche de mots vedettes dans un ou plusieurs dictionnaires) comme le prouvent les réalisations des projets les plus connus *Ajeeb* du système *Sakhr* [<http://lexicons.ajeep.com/Results.asp>], *Alburaq* [<http://www.alburaq.net/>], *Kalimet* [<http://www.kl28.com/lesanalarab.php>] et *LisanElArab/Al-Qamous-Al-Mouhit* [<http://www.content.com.sa/Languages/>]. Ces limites d'interrogation sont dues, principalement, à une faiblesse de structuration des entrées dictionnaires utilisées.

Devant ces limites, et en vue de donner à cette langue la place qui lui convient en tant que langue véhiculaire, nous proposons une modélisation LMF pour l'élaboration de dictionnaires arabes éditoriaux extensibles et incrémentaux (possibilité d'ajouter des entrées et des descriptions au sein des entrées). Cette modélisation permettra de profiter de la richesse de la langue arabe en unifiant ce qui existe avec une structure évolutive et aussi fine que possible, à

partir de laquelle, il est possible de réaliser des fonctions de consultation génériques et adaptées aux besoins variés des utilisateurs.

Dans cet article, nous commençons par présenter les principales caractéristiques des dictionnaires de l'arabe, en insistant notamment sur leurs spécificités structurelles. Ensuite nous argumentons le choix de la modélisation LMF pour l'élaboration de dictionnaires électroniques arabes à usage humain. Puis, nous détaillons le modèle normalisé que nous proposons. Enfin, nous décrivons une mise en œuvre informatique et une expérimentation du modèle partant de la normalisation du dictionnaire éditorial *El-Ghani* « الغني » et d'un ensemble d'extraits relatifs à d'autres dictionnaires. L'interrogation est réalisée par le système ADIQT0 (Arabic DIctionary Query TOols) que nous avons développé pour l'interrogation générique des dictionnaires normalisés de l'arabe.

2 Principales caractéristiques des dictionnaires arabes

La lexicographie arabe est une discipline très ancienne. Elle a connu tout le long de son histoire une diversité d'écoles ayant chacune ses spécificités. Certaines de ces écoles proposent une classification des entrées lexicales selon l'ordre phonétique (e.g. *Al-Ain* « العين », *Tahdhib Al-Loghah* « تهذيب اللغة »). D'autres utilisent l'ordre alphabétique soit direct (e.g. *Lisan-El-Arab* « لسان العرب », *Al-Qamous-Al-Mouhit* « القاموس المحيط », *Assihah* « الصّحاح ») soit inversé (e.g. *El-Ghani* « الغني », *Larousse : Al-Moojam Alarabi Al-Hadith* « لاروس: المعجم العربي الحديث »). Nous distinguons aussi le classement selon l'ordre alphabétique croissant de la racine du mot qui constitue l'une des spécificités des dictionnaires de l'arabe (e.g. *Al-Wassit* « الوسيط », *Al-Mosbah Al-Mounir* « المصباح المنير »). Cette organisation permet d'inclure toutes les formes verbales et nominales dérivées d'une racine sous la même entrée.

Cette diversité peut être résumée dans les points suivants :

- Contenu : la description du contenu des entrées varie d'un dictionnaire à un autre. Certains se contentent des exemples pour définir le sens d'un mot, d'autres donnent la définition accompagnée d'exemples. Les informations d'ordre morphologique et syntaxique sont plus ou moins riches d'un dictionnaire à un autre, etc.
- Macrostructure : l'organisation des entrées lexicales diffère d'un dictionnaire à un autre (selon l'école lexicographique). Par exemple, *MaKTaBaTun* « مكتبة » (librairie) représente une entrée lexicale à part dans le dictionnaire *El-Ghani* « الغني » alors qu'il figure comme une sous-entrée de l'entrée lexicale *KaTaBa* « كَتَبَ » (écrire) dans le dictionnaire *Al-Wassit* « الوسيط » puisque c'est un mot dérivé de *KaTaBa*.
- Microstructure : l'organisation des informations linguistiques, au niveau des entrées lexicales, varie d'un dictionnaire à un autre et même au sein d'un même dictionnaire. Par exemple, le pluriel et/ou le féminin d'un lemme de type nom "اسم" ou adjectif "صفة" se trouvent en tête de l'article dans le dictionnaire *El-Ghani* « الغني » et à la fin de l'article, si elles existent, dans *Al-Wassit* « الوسيط ».
- Consultation : Il n'y a pas une manière unique de consulter ces dictionnaires. Dans la plupart des cas, l'utilisateur est contraint à connaître la structure interne d'un dictionnaire pour pouvoir trouver l'information qu'il cherche.

La question que nous nous posons est comment profiter de cette richesse structurelle et informationnelle dans l'élaboration de nouveaux dictionnaires électroniques arabe ?

3 Choix et présentation générale de LMF

3.1 Choix de LMF

En vue de proposer un modèle unifié des dictionnaires arabes à usage éditorial, notre choix s'est fixé sur LMF pour les raisons suivantes. Tout d'abord, cette future norme, qui sera publiée très prochainement par l'ISO sous le numéro 24613, permet la spécification de lexiques monolingues et multilingues destinés à la fois aux usages éditorial et TALN. En plus, elle assure une modélisation extensible et modulaire couvrant tous les niveaux de description linguistique (i.e. morphologie, syntaxe, sémantique). De surcroît, sa souplesse de modélisation favorise la représentation des caractéristiques de la langue arabe notamment ses aspects dérivationnel et flexionnel (Khemakhem, Gargouri, Abdelwahed & Francopoulo, 2007). En effet, les classes des extensions de LMF permettent de représenter la racine et le schème (les schèmes sont des modèles de différentes structures que l'on applique aux racines) d'une entrée lexicale du dictionnaire, ainsi que les patrons morphologiques assurant les opérations de dérivation et de flexion. De ce fait, il est possible de représenter aussi bien les lexiques extensionnels qu'intensionnels. Enfin, la modélisation LMF se prête à une utilisation en ligne avec les services Web (Ben Abderrahman, Chaari, Gargouri & Jmaiel, 2006).

3.2 Présentation générale de LMF

Dans une perspective de généricité, LMF propose un méta-modèle constitué d'une partie noyau obligatoire et de modèles d'extension optionnels couvrant, notamment, les aspects morphologique, syntaxique, sémantique et MRD (Machine Readable Dictionary) (Francopoulo & George, 2008). La modularité de LMF est assurée par les mécanismes d'adjonction des catégories de données aux composantes du méta-modèle. Ce sont des descripteurs linguistiques élémentaires qui permettent de décorer les classes du méta-modèle par les spécificités de la langue et du lexique (e.g. /Part of speech/, /Grammatical number/, etc.). Elles sont organisées, avec leurs valeurs possibles, dans un registre de catégories de données qui est consultable et éditable en ligne [<http://syntax.inist.fr>]. Ainsi leur normalisation suit des principes définis par la norme ISO 12620 (Ide & Romary, 2004).

4 Elaboration du modèle normalisé des dictionnaires arabes

Ce modèle propose une représentation unifiée des dictionnaires arabes existants en se référant à la dernière révision de LMF (rev.15) (Francopoulo & George, 2008).

4.1 Démarche de création du modèle normalisé

Dans ce paragraphe, nous nous proposons d'esquisser un modèle normatif et représentatif des dictionnaires arabes. La démarche suivie respecte le processus recommandé par LMF (Francopoulo & George, 2008). Ainsi, nous avons commencé par la sélection des classes nécessaires pour représenter les informations dictionnairiques de l'arabe. Les classes que nous avons retenues représentent pour notre modèle cible des sous modèles relatifs aux extensions morphologique, MRD, syntaxique et sémantique. Ces classes sont greffées sur le noyau de LMF qui spécifie essentiellement les notions de lexique, de mot, de forme et de sens. La présence du noyau est obligatoire selon LMF. Ensuite, nous avons sélectionné les catégories

de données adéquates à l'arabe à partir du registre normalisé. Enfin, nous avons décoré les sous modèles ainsi créés par les catégories de données sélectionnées.

4.2 Modèle de l'extension morphologique

Le modèle de cette extension permet de représenter les informations morphologiques des lexiques. Compte tenu de l'aspect flexionnel et dérivationnel de l'arabe, nous avons retenu les classes *Lemma*, *Related Form* et *Word Form* de l'extension morphologique de LMF, qui servent à décrire, respectivement, le schème et la forme écrite du lemme, sa racine et les propriétés des formes fléchies. La Figure 1-a décrit les liens des classes morphologiques avec celles du noyau. Elle précise, en plus, les catégories de données utilisées. La Figure 1-b présente un extrait XML qui illustre l'instanciation de ce modèle pour l'exemple de l'entrée lexicale *KaTaBa* « كَتَبَ » (*écrire*).

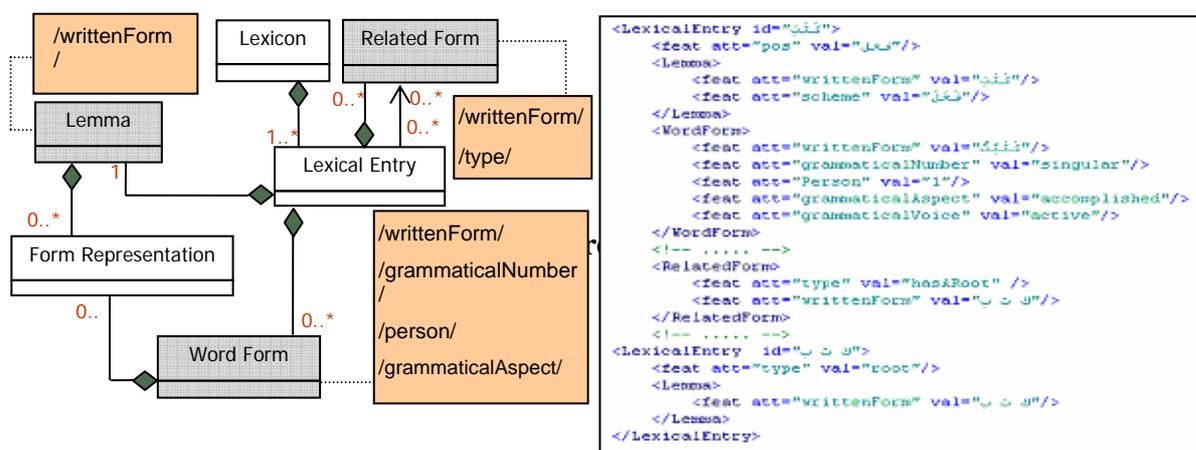


Figure 1-a : Modèle morphologique Figure 1-b : Extrait XML du modèle morphologique

4.3 Modèle de l'extension MRD

Cette extension est particulièrement utilisée dans la description des dictionnaires électroniques à usage éditorial. La Figure 2 présente les liens des classes MRD, que nous avons retenues, *Context* et *SubjectField* avec la classe *Sense* du noyau. La première classe représente un exemple d'utilisation du lemme dans une phrase. Quant à la deuxième, elle définit le domaine d'usage du lemme correspondant à un sens donné. Nous avons exclu la troisième classe MRD, proposée par LMF, à savoir *Equivalent*, vu que cette dernière est relative à un contexte multilingue.

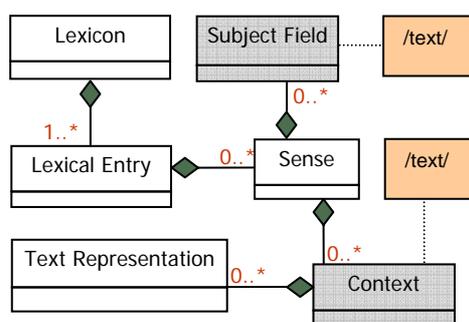


Figure 2 : Modèle de l'extension MRD

4.4 Modèle de l'extension syntaxique

L'objectif de ce modèle est de décrire les propriétés syntaxiques du mot vedette lorsqu'il s'agit d'une construction formée de plusieurs mots. Les classes que nous jugeons nécessaires pour décrire les informations syntaxiques de l'arabe sont *Syntactic Behaviour*, *Subcategorization Frame* et *Subcategorization Frame Set* qui sont rattachées au noyau comme le montre la Figure 3-a. Le rôle de ces classes correspond bien aux descriptions d'ordre syntaxique utilisées dans la majorité des dictionnaires de l'arabe là où la focalisation sur un verbe suggère la description de sa valence. La Figure 3-b, donnée ci-après, illustre un extrait XML d'une instance du modèle syntaxique relative au mot *KaTaBa* « كَتَبَ » (*écrire*).

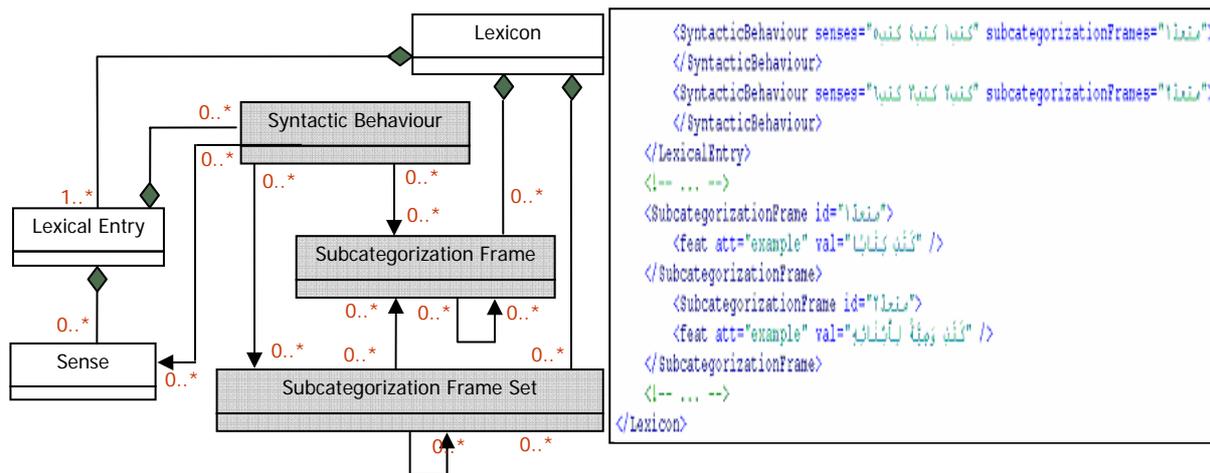


Figure 3-a : Modèle syntaxique

```

<SyntacticBehaviour senses="كتب( كُتِبَ) كَتَبَ" subcategorizationFrames="منعاً">
</SyntacticBehaviour>
<SyntacticBehaviour senses="كتبا كُتِبَ كَتَبَ" subcategorizationFrames="منعاً">
</SyntacticBehaviour>
</LexicalEntry>
<!-- ... -->
<SubcategorizationFrame id="منعاً">
<feat att="example" val="كُتِبَ كَتَبَ" />
</SubcategorizationFrame>
<SubcategorizationFrame id="منعاً">
<feat att="example" val="كُتِبَ وَجِبًا بِأَنْتَابِهِ" />
</SubcategorizationFrame>
<!-- ... -->
</Lexicon>
    
```

Figure 3-b : Extrait XML du modèle syntaxique

4.5 Modèle de l'extension sémantique

En ce qui concerne l'extension sémantique de notre modèle, nous avons retenu les classes *SenseRelation* et *SenseExample*. Le choix de la classe *SenseRelation*, à la place de *Synset*, proposée aussi par LMF, se justifie par le fait que nous ne disposons pas de dictionnaires type WordNet qui se base sur le concept de synset (ensemble de synonymes). La classe *SenseExample* s'avère très utile pour la langue arabe. En effet, la totalité des dictionnaires de cette langue insistent sur les exemples, notamment pour définir les mots vedettes.

Les deux classes que nous avons retenues sont reliées à la classe *Sense* du noyau comme le montre la Figure 4-a. Par ailleurs, la Figure 4-b illustre un extrait d'instanciation de ce modèle, décrit en XML.

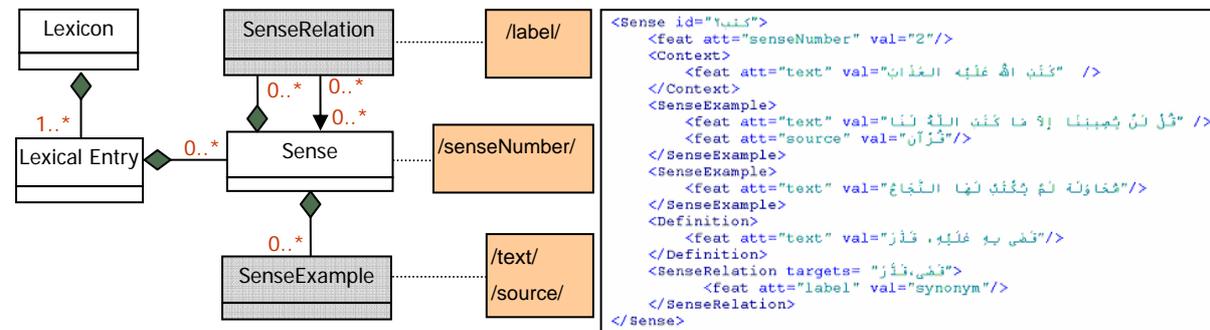


Figure 4-a : Modèle sémantique

```

<Sense id="كتبا">
<feat att="senseNumber" val="2"/>
<Context>
<feat att="text" val="كُتِبَ اللهُ عَلَيْهِ الْعَذَابُ" />
</Context>
<SenseExample>
<feat att="text" val="قُلْ لَنْ يُغَيِّبَنَا إِذْ مَا كُنْتَ إِلَيْهِ لِنَا" />
<feat att="source" val="قُرْآن"/>
</SenseExample>
<SenseExample>
<feat att="text" val="مُخَاوَنَةٌ لَمْ يُغَيِّبْنَا لَهَا التَّجَاعُ" />
</SenseExample>
<Definition>
<feat att="text" val="تَغَيَّبَ بِهِ عَلَيْهِ، قَذَرَ" />
</Definition>
<SenseRelation targets="تَغَيَّبَ، قَذَرَ">
<feat att="label" val="synonym"/>
</SenseRelation>
</Sense>
    
```

Figure 4-b : Extrait XML du modèle sémantique

5 Mise en œuvre du modèle et expérimentation

Dans le but d'expérimenter le modèle élaboré, nous avons tout d'abord procédé à la normalisation de certains dictionnaires disponibles, notamment *El-Ghani* « الغني ». Nous avons choisi, dans ce cas, une représentation extensionnelle de ce dictionnaire puisqu'il ne comporte pas toutes les formes fléchies d'un mot. Ensuite, nous avons défini un ensemble de fonctions d'interrogation selon le niveau des utilisateurs. Enfin, nous avons développé et testé ces fonctions dans le cadre de la mise en œuvre du système d'interrogation ADIQTO.

5.1 Normalisation des dictionnaires existants

Nous avons procédé, en premier lieu, à la normalisation du dictionnaire *El-Ghani* « الغني » dont nous disposons d'une version HTML. En effet, ce dictionnaire se distingue par sa microstructure allégée et invariante, ainsi que par un ensemble de marqueurs facilitant la transformation de son contenu brut vers une version structurée en XML. Par ailleurs, cette conversion a été faite d'une manière quasi-automatique en se servant d'outils TALN disponibles dans notre laboratoire (i.e., segmenteur, analyseur morphologique).

En plus, nous avons normalisé quelques extraits d'autres dictionnaires *Al-Wassit* « الوسيط », *Lissan-El-Arab* « لسان العرب » et *Al-Mouhit* « المحيط » dont la structure est plus complexe, qui varie d'une entrée à une autre et caractérisée par une quasi-absence de marqueurs. Cette normalisation a été réalisée manuellement vu la difficulté de la mise en place d'un système automatique d'analyse et de conversion.

5.2 Elaboration des fonctions d'interrogation

Pour définir les fonctionnalités d'interrogation des dictionnaires, nous avons considéré deux critères : l'intervention de l'utilisateur et la complexité de la requête. Ainsi, les services d'interrogation peuvent être classés en trois types : (i) interrogation simple, (ii) interrogation guidée ou assistée et (iii) interrogation avancée (complexe). La recherche peut s'effectuer dans un ou plusieurs dictionnaires choisis par l'utilisateur.

- Interrogation simple

Cette recherche permet un accès simple à un mot vedette (entrée lexicale principale) ou à l'une de ses formes fléchies. Toutes les informations relatives au mot recherché seront affichées : caractéristiques morphologiques, définitions, exemples, etc.

- Interrogation guidée ou assistée

L'utilisateur de ce type de recherche jouit d'une assistance à travers une correction du mot saisi. De surcroît, des requêtes prédéfinies seront mises à la disposition consistant à la recherche des synonymes, du masculin ou du féminin, du singulier ou du pluriel et finalement des verbes et des noms dérivés.

- Interrogation avancée

Le troisième type concerne des besoins complexes exprimés par des experts qui cherchent à calculer des statistiques sur la morphologie des mots, les liens entre les mots, etc. ou encore à

calculer des résultats spécifiques non stockés dans le dictionnaire comme la conjugaison. Dans ce cadre, nous allons réutiliser le langage graphique LMF-QL (Ben Abderrahman, Gargouri & Jmaaiel, 2007). LMF-QL permet de générer une requête XQuery à partir d'une sélection dynamique des entrées qui seront combinées avec des opérateurs logiques (*not*, *and* et *or*). Ainsi, l'utilisateur est capable d'accéder à n'importe quelle information et à faire des croisements selon le besoin envisagé (e.g. chercher les mots qui ont la même racine).

5.3 Implémentation du système ADIQTO

L'implémentation du système ADIQTO est réalisée en Java avec utilisation de l'API Saxon9 pour assurer l'exécution des requêtes XQuery. Cette implémentation se base sur des requêtes prédéfinies assurant, notamment, les recherches simples et assistées (des macros). Quant à la recherche avancée, ses requêtes seront générées automatiquement grâce à l'utilisation du langage LMF-QL. La Figure 5 illustre la recherche simple du mot *KaTaBa* « كَتَبَ » (*écrire*).



Figure 5 : Interface graphique de recherche simple du système ADIQTO

Outre les services de consultation, ADIQTO assure aussi bien la correction des erreurs de voyellation du mot recherché que celle de ses erreurs orthographiques par le biais du système CORTEXA (Ben Hamadou, 1993) qui, dans un premier temps, analyse les mots après dévoyellation, puis propose une liste de mots candidats à la correction du mot recherché.

6 Discussion

Le système ADIQTO est actuellement fonctionnel sur le dictionnaire moderne *El-Ghani* « الغني » et sur des fragments des dictionnaires : *Al-Wassit* « الوسيط », *Al-Mouhit* « المحيط » et *Lissan-El-Arab* « لسان العرب ». Le modèle que nous avons proposé a élargi, comme prévu, le champ d'accès aux données dictionnaires, en permettant des requêtes très variées afin de satisfaire les besoins d'une diversité d'utilisateurs. A l'inverse des outils de consultation des dictionnaires arabes disponibles, le principal mérite d'ADIQTO consiste à fournir un accès sélectif aux différentes informations relatives à une entrée lexicale. A cet égard, il n'est plus obligatoire de limiter l'interrogation à la liste des entrées lexicales, mais elle peut couvrir

n'importe quelle forme fléchie ou dérivée rattachée à une entrée lexicale. D'un autre côté, le modèle ainsi défini supporte les représentations extensionnelle et intentionnelle, ce qui permet de produire des dictionnaires complets avec génération automatique des formes fléchies.

7 Conclusion

Dans cet article, nous avons proposé un modèle conforme à la norme LMF-ISO 24613 pour la représentation des dictionnaires électroniques éditoriaux de la langue arabe. Ce modèle a le mérite d'unifier les différents modèles proposés en lexicographie de l'arabe tout en offrant des possibilités d'extension. La structure évolutive et assez fine de ce modèle a permis un accès sélectif aux différents champs afférents à un mot vedette par le biais d'un ensemble de recherches (simple, assistée et avancée) satisfaisant les besoins d'une diversité d'utilisateurs (i.e. apprenant de la langue, écrivain, journaliste, linguiste. etc.).

Dans les travaux futurs, nous comptons développer des services Web permettant de mettre en ligne le système ADIQTO, muni des dictionnaires normalisés disponibles, en vue d'élargir le spectre de son utilisation. Finalement, nous envisageons aborder la normalisation des dictionnaires éditoriaux existants de l'arabe, notamment à partir des versions numériques dont nous disposons (nous disposons d'une dizaine de dictionnaires en format .doc, html, et pdf).

Références

- AIT TALEB S., (2005). Dictionnaires électroniques arabes : le modèle des dictionnaires de Sakhr, *revue de l'Association Marocaine des Etudes Lexicographiques*, Numéro 3-4, 15-31.
- ARREGI X. & AL. (2002). Semiautomatic of the Euskal Hiztegia Basque Dictionary to a queryable electronic form. *L'objet* 8/2002, LMO' 2002, 45-57.
- BEN ABDERRAHMEN M., CHAARI F., GARGOURI B., JMAIEL M. (2006). Des services orientés besoin pour l'exploitation des bases lexicales normalisées. *10th Maghrebien Conference on Software Engineering and Artificial Intelligence MCSEAI'06, Agadir, Maroc.*
- BEN ABDERRAHMEN M., GARGOURI B., JMAIEL M. (2007). LMF-QL: A graphical Tool to Query LMF databases. *Third Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland.*
- BEN HAMADOU A., (1993). Vérification et correction automatiques par analyse affixale des textes écrits en langage naturel : le cas de l'arabe non voyellé. *Thèse Es-Sciences en Informatique, Faculté des Sciences, Université de Tunis.*
- DENDIEN, J. & PIERREL, J.-M. (2003). Le Trésor de la Langue Française informatisé. Un exemple d'informatisation d'un dictionnaire de langue de référence. *TAL* 44, Numéro 2, 11-39.
- FRANCOPOULO G. (2004). Proposition de norme des lexiques pour le traitement automatique du langage. INRIA/LORIA-ACTION SYNTAXE, Version-1.10, 13 mai 2004.
- FRANCOPOULO G., GEORGE M. (2008). ISO/TC 37/SC 4 Rev.15. Language resource management – Lexical markup framework (LMF).

IDE N., ROMARY L. (2004). International standard for a linguistic annotation framework. *International Journal of Natural Language Engineering*, 10 Numéro 3-4, 211-225.

KHEMAKHEM A., GARGOURI B., ABDELWAHED A., FRANCOPOULO G (2007). Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF - ISO 24613. *Traitement Automatique des Langues Naturelles : du 5 au 8 juin 2007 à Toulouse*.

SPERBERG-MCQUEEN C.M., BURNARD L. (2005). TEI P5 – Guidelines for Electronic Text Encoding and Interchange, TEI Consortium. January 2005.

WOOLDRIDGE R. (1998). L'informatisation du Dictionnaire de l'Académie française (DAF). DictA 1998, Table ronde sur l'informatisation des dictionnaires anciens. Limoges, 19-20.