Al Qamus al Muhit, a Medieval Arabic Lexicon in LMF

Ouafae Nahli, Francesca Frontini, Monica Monachini
Istituto di Linguistica Computazionale "A. Zampolli", CNR, Pisa
ouafae.nahli@ilc.cnr.it, francesca.frontini@ilc.cnr.it, monica.monachini@ilc.cnr.it

Fahad Khan
Dipartimento di Studi Umanistici, Università di Ca' Foscari, Venezia
Istituto di Linguistica Computazionale "A. Zampolli", CNR, Pisa
fahad.khan@unive.it, fahad.khan@ilc.cnr.it

Arsalane Zarghili, Mustapha Khalfi
Faculté des Sciences et Techniques, Université Sidi Mohamed Ben Abdellah, Fez
zargili@gmail.com, mus.khalfi@gmail.com

Abstract

This paper describes the conversion into LMF, a standard lexicographic digital format of *'al-qāmūs al-muḥīṭ*, a Medieval Arabic lexicon. The lexicon is first described, then all the steps required for the conversion are illustrated. The work is will produce a useful lexicographic resource for Arabic NLP, but is also interesting per se, to study the implications of adapting the LMF model to the Arabic language. Some reflections are offered as to the status of roots with respect to previously suggested representations. In particular, roots are, in our opinion are to be not treated as lexical entries, but modeled as lexical metadata for classifying and identifying lexical entries. In this manner, each root connects all entries that are derived from it.

Keywords: Arabic Lexicon, LMF, Al Qamus al Muhit

1. Introduction

In this article we will describe ongoing work in the conversion of a medieval Arabic lexicon *'al-qāmūs al-muḥīṭ* into the XML format using the Lexical Markup Framework (LMF). Our overall aim is to make this historically important lexicographic work available in a format that renders the lexical information contained within it more easily accessible and especially by NLP applications. The utility of LMF in this case is that it provides a standardized framework for modeling and representing NLP lexicons as has been well demonstrated in several previous projects (Francopoulo 2013).

In the next section we will discuss the different challenges that arise due to the structure of the Arabic language itself and which it is necessary to take into consideration when processing Arabic lexicons which are generally structured quite differently from lexicons for European languages like English or Italian. We will briefly detail some background on the *'al-qāmūs al-muḥīṭ* lexicon itself and outline the particular mode of organisation used in that work. We will also describe previous work that has been carried out in developing Arabic language lexica using the LMF format. In the following section we will present the process by which the original text file used as a source for the lexicon was processed and segmented to extract structured information from the original entries and definitions, and how this information was then represented in LMF compliant entries. In doing so, we shall briefly address some open issues with respect to the correct representation of Arabic roots as well as looking forward to our plans for future work.

2. Background

Words in Arabic as in other Semitic languages like Hebrew and Aramaic, are formed in a systematic way on the basis of consonantal roots. This systematicity in word formation gives us a natural means of categorising and grouping together the lexical entries in the language. For instance, *kataba* ("he writes"), *maktab* ("office"), *maktabah* ("library") *'istaktaba* ("he wrote for himself") and *kitāb* (book) are separate lexical entries that are all derived from the same root: that is, morphologically they all share the same three consonants that together constitute the root *ktb*. The order of the radical consonants is fixed and therefore we can also speak of the consonantal skeleton. More precisely, Arabic words are formed by the attribution of a "scheme" (consisting of vowels and derivational affixes) to a root (Moutaouakil, 1989, p. 13). For example, the verb *kataba* "to write is obtained by attributing the verbal scheme *fa'ala* to the root *ktb*, the noun *maktab* "desk is formed attributing the scheme *maf'al* in which the prefix *ma* indicates the event location. Therefore each unit is defined by two criteria: a) a morpho-semantic criterion, namely, the root and ii) a morpho-syntactic criterion namely the scheme (Kouloughli, 1994). On the one hand, words that convey a similar semantic meaning have the same root. On the other hand, all the words that convey the same "grammatical function" have the same form: the scheme (Cantineau, 1950, Dichy 2002).

This property of Arabic makes it extremely useful to be able to classify words using consonantal roots and schemas since it enables us to capture important kinds of semantically relevant information, e.g., in the example above, the semantic link that unites *maktab*, *maktabah*, and *kitāb* is the fact that they are, respectively, the event, place and object of the action “write”. In order to make this information more accessible, lexical entries in (printed) Arabic dictionaries are usually classified according to the root from which they are derived rather than on the basis of the first character of the entry as is usually the case with languages like English or Italian. The proper treatment of roots and schemes is also important in computational lexicons as we shall later see.

2.1. The source text

We made the decision to work with the Arabic lexicon *’al-qāmūs al-muḥīṭ* (AQAM) for a number of reasons but primarily because of to its authoritative status in the Arabic speaking world and its comprehensiveness in terms of the number of entries contained within it. The original compiler of AQAM, the medieval author ’al-fīrūz’ābādī (1329-1414), states in his introduction to the work that it was created by merging together several pre-existing dictionaries. For this reason, he gave it the title *’al-qāmūs* (the Ocean) *al-muḥīṭ* (Universal). AQAM is widely regarded as the most authoritative lexicon of Arabic ever published. Indeed the word *qāmūs* has even come to supplant the word *munǧid* (dictionary). As part of the process of compilation, ’al-fīrūz’ābādī greatly reduced the original content from the source dictionaries he was using by eliminating examples, Quranic quotations, poetry and some grammatical information. In the end the fact that AQAM is well structured as a lexicon and the fact that it contains short lexical items make it an excellent candidate for conversion into a computational lexical.

Traditionally there have been several schools of thought within Arabic lexicography with respect to the ordering of consonantal roots in the lexicon (Carter, 1990), (Lancioni 1997). AQAM classifies entries initially by the final consonant of the root; this gives an organisation of the (printed) lexicon into 28 chapters. Each chapter is then further organised into sub-chapters according to the first consonant of the entries contained within. These sub chapters are organised in their turn according to the second consonant of the root[1].

3. Processing the text

3.1. First steps

Our original source was a digitized version of AQAM chosen for its accuracy after comparing the different available online versions with an authoritative paper edition [2]. The original version of AQAM was in .txt format and structured as in Fig. 1.

To normalize the text of AQAM in this first electronic version, and as a first step towards converting it into a more usable format, we tried to identify the lexicographic conventions used to organise the information in the text markers, its meta-structure. This was carried out manually. We were able to identify both explicit indicators, such as *1* and *2* indicating respectively the third and the first radical, and implicit indicators such for example, the symbol @ which indicates the change of the second radical without necessarily allowing its identification. We were finally therefore able to segment the text and identify each lexical entry with its definition as in Fig 2.

Here the commas and the colon serve to separate the different pieces of information: lexical information on the one hand, and morphological information on the other. This can in most part be automatically derived from some basic background knowledge about Arabic grammar. In Figure 2, we present the example of the verb *kabura* which is equivalent in meaning to the English predicate “grow; to be great”. In this instance the main lexical entry (*kabura*) is followed, after a comma separator, by its related masḍars (in the accusative indefinite form). Lexical information is placed after the colon, followed by other relevant information.

Kabura belongs to the *fa’ula*[3] scheme and thus to the verbal model that expresses quality or status; all verbs in this class are accompanied by a related adjective that belongs to the *fa’īl* scheme. In the AQAM, the adjective is always introduced by *fa=huwa* “therefore it/he is” and followed by morphosyntactic information regarding the feminine case and possible broken plurals. As this example shows there is a lot of implicit information in each entry which we can use to determine the most important properties of the lexical entry in question; this process can in many cases be automated.

In this first phase of the treatment of the text, the segmentation is coarse: each block of extracted text represents a separate entry in the lexicon and is tagged as “Unsegmented Text”. In the next phase, the resulting text is divided into definitions and lemmata. This latter can consist of a single lemma or a set of lemmata that must be identified. An example of the formatted text that results from the application of these two phases can be seen in Fig. 3.

3.2. The LMF standard

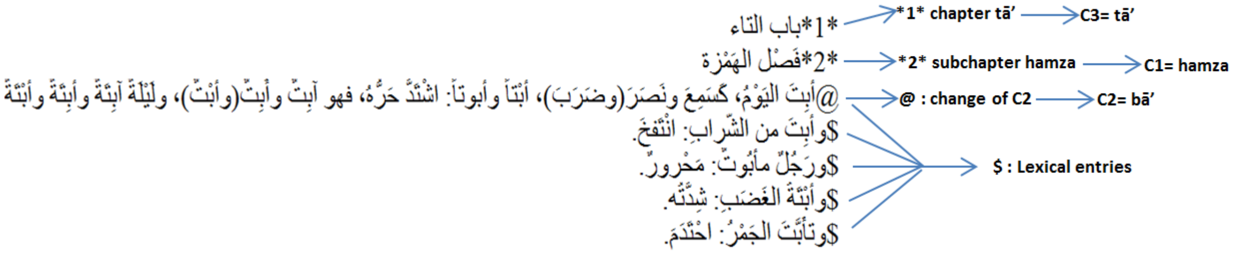
LMF is a model for representing computational lexicons that has the status of an ISO Standard. It was developed as a joint effort by a team of specialists in computational lexicography and natural language processing and has been used in a number of lexicographic projects, and as the input format in several NLP applications and tools (Del Grosso et al. 2014). The applicability of LMF to Arabic language resources has been demonstrated in several works, such as (Khemakhem et al. 2007, 2009, 2013; Loukil 2007, Attia et al 2010).

We chose to represent AQAM in LMF so as to make the information contained within the lexicon available in a structured form both for lexicographic research and for NLP applications. Thus not only will it be possible for researchers

[1] Encyclopedia of Arabic Literature (1998), Volume 2, eds. Julie Scott Meisami, Paul Starkey, Routledge 1998.

[2] ’al-fīrūz’ābādī, ’al-qāmūs al-muḥīṭ, (ed.) al- arqsūsī M. N. muassasat ar-risālah, 1998. Beirut.

[3] The schema is a syllabic pattern in which the root consonants (R) occupy a specified place; derivational affixes may be inserted at specified positions; but also the length and tone of vowels are specified. In Arabic, the formal representation of the schema is done using the consonants of the root *f’l* from which derives the verb *fa’ala* “do”. Consonants *f’l* respectively represent the first, the second and the third radical of any root. For example the scheme *fa’ula* represents simply the “form R1aR2uR3a and *fa’l* is R1aR2R3 (Dichy 2002).



*1*باب التاء

*2*فَصْلُ الهَمْزة

@أبِتَ اليَوْمُ، كَسَمِعَ ونَصَرَ(وضَرَبَ)، أبْتاً وأبوتاً: اشْتَدَّ حَرُّهُ، فهو آبِتٌ وأبِتٌ(وأبْتٌ)، ولَيْلَةٌ آبِتَةٌ وأبِتَةٌ وأبْتَةٌ

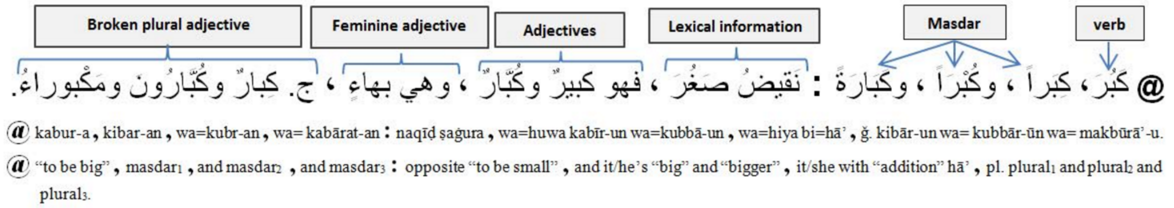
$وأبِتَ من الشَّرابِ: انْتَفَخَ.

$ورَجُلٌ مأبُوتٌ: مَحْرورٌ.

$وأبْتَةُ الغَضَبِ: شِدَّتُه.

$وتأبَّتَ الجَمْرُ: احْتَدَمَ.

Figure 1: The structure of AQAM explained

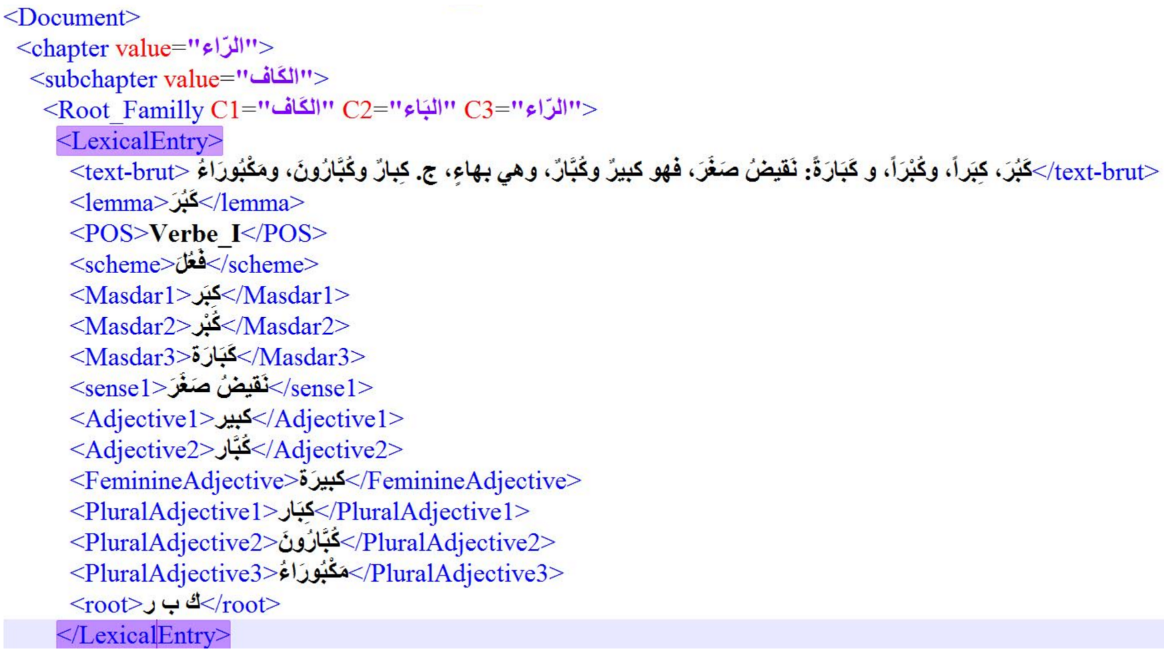


@كَبُرَ، كِبَراً ، وكُبْراً ، وكَبَارَةً : نَقِيضُ صَغُرَ ، فهو كبيرٌ وكُبَّارٌ ، وهي بهاءٍ ، ج. كِبارٌ وكُبَّارُونَ ومَكْبوراءُ.

@ kabur-a , kibar-an , wa=kubr-an , wa= kabārat-an : naqīḍ ṣaġura , wa=huwa kabīr-un wa=kubbā-un , wa=hiya bi=hā’ , ğ. kibār-un wa= kubbār-ūn wa= makbūrā’-u.

@ “to be big” , masdar$_1$, and masdar$_2$, and masdar$_3$: opposite “to be small” , and it/he’s “big” and “bigger” , it/she with “addition” hā’ , pl. plural$_1$ and plural$_2$ and plural$_3$.

Figure 2: An entry of the digitized AQAM explained.



```
<Document>
 <chapter value="الرّاء">
  <subchapter value="الكَاف">
   <Root_Familly C1="الكَاف" C2="البَاء" C3="الرّاء">
    <LexicalEntry>
     <text-brut> كَبُرَ، كِبَراً، وكُبْرَاً، و كَبَارَةً: نَقيضُ صَغُرَ، فهو كبيرٌ وكُبَّارٌ، وهي بهاءٍ، ج. كِبارٌ وكُبَّارُونَ، ومَكْبُورَاءُ</text-brut>
     <lemma>كَبُرَ</lemma>
     <POS>Verbe_I</POS>
     <scheme>فَعُلَ</scheme>
     <Masdar1>كِبَر</Masdar1>
     <Masdar2>كُبْر</Masdar2>
     <Masdar3>كَبَارَة</Masdar3>
     <sense1>نَقيضُ صَغُرَ</sense1>
     <Adjective1>كبير</Adjective1>
     <Adjective2>كُبَّار</Adjective2>
     <FeminineAdjective>كبيرَة</FeminineAdjective>
     <PluralAdjective1>كِبَار</PluralAdjective1>
     <PluralAdjective2>كُبَّارُونَ</PluralAdjective2>
     <PluralAdjective3>مَكْبُورَاءُ</PluralAdjective3>
     <root>ك ب ر</root>
    </LexicalEntry>
```

Figure 3: Initial segmentation of the digitized text.

to query and search for specific classes of words and see how they are described by AQAM, but also to use this information for the processing of Arabic texts, in particular old texts.

We decided to use the NLP module of LMF in its canonical form. In order to do this, the information contained in the AQAM and made explicit in the initial segmentation (see Figure 3) had to be fitted into the LMF constraints. We based our initial attempts to convert AQAM into XML using LMF on the examples given in (Khemakhem 2013). Below we give a brief description of the encoding along with an example.

First of all a lexical resource of type lexicon is instantiated. It contains information about the language and the lexicon as well as all the metadata concerning the source of information. Then the lexical entries are listed. In Figure 4 below we represent the LMF encoding of the verb *kabura* which was discussed above. In this case a single entry of the AQAM generates several lexical entries in LMF. Not only does the verb need to be represented, but also its related forms (*masdar*, adjectives, ...). In addition relations between lexical entries and their roots are made explicit. Finally different senses are created for each lexical entry based on the various definitions (although note that we havent included all the resulting lexical entries and relations between entries in the diagram for reasons of clarity). As we mentioned above, so far the examples which we have encoded in LMF have been carried out according to the examples given in (Khemakhem 2013). This has entailed encoding roots as Lexical Entries with type root (see LE kbr in the figure). However, the choice of encoding roots as LE is somewhat problematic for theoretical and practical reasons. First of all the definition of LE in LMF implies that a LE should be lexeme, namely a unit with precise lexical meaning, which doesn't seem to be applicable to roots. Arabic roots, in fact, group lexemes in semantic classes, but are not themselves provided with a specific meaning. As a consequence of this by adding roots as lexical entries lexicographers are promoting what is just a device to group semantically related words in printed dictionaries into actual words, thus inflating the amount of "real" entries in the actual lexicon. This doesn't mean that roots shouldn't be listed as entries in the lexicon: in fact the regularity of derivation in Arabic is such, that anyone using an Arabic lexicon expects to be able to search for a root and extract all lexical entries bearing that root. For this reason a viable alternative would be to change the proposal in Figure 4 in such a way as to instantiate Arabic roots as a separate class of objects, such as "Root" distinct from that of Lexical Entry. In order to do this, we intend to propose the insertion of such a category in the registry, so that it can be legitimately used in LMF. A similar treatment might be envisaged for schemes.

4. Commented example

In the Appendix to this paper we present a full LMF example of the lexical entry *kabura* . As we mentioned above, *kabura* appertains to the *fa'ula* scheme. Formally, all verbs of the perfective scheme *fa'ula* (R1aR2uR3a) have the imperfective scheme *yaf'ulu* (yaR1R2uR3u). Many morpho-syntactic and morpho-semantic features of verb depend on the second vowel. For example, in this case, the second vowel is /u/ in both the perfective and impefective forms, and the verb belongs of the class (u/u). Understandably all verbs of this morphological class (u/u) are stative verbs, therefore, they are intransitive verbs and are connected to an adjective describing this state. The scheme formalism makes it possible to generate a lot of information automatically: inflectional paradigm (u/u)[4]; some syntactic features. In addition, other lexical entries (for each masdar and adjective) can be created and enriched by additional information automatically.

In the Appendix example we also present some links to external resources, implemented using the Monolingual External Ref component. This component is used to link a monolingual lexicon (such as ours) to an external resource that may be also in a different language[5]. In particular two Monolingual External Refs are created, one to link the first sense of *kabura* to a SUMO (Pease et al., 2002) ontological node and one to link the same sense to a Princeton Wordnet 3.1 (Fellbaum, 1998) synset.

5. Conclusion and future work

We plan to deliver the chapter of the letter *bā'* by the time of LREC 2016. Future work will concern the complete mapping of this lexicon to other lexical and conceptual; in addition to the aforementioned Princeton Wordnet 3.1 and SUMO, we plan to link entries to resources for Arabic, such as Arabic WordNet (Black et al., 2006; Rodríguez et al., 2008a; Rodríguez et al., 2008b; Elkateb et al., 2006).

6. Acknowledgements

We thank Angelo Mario Del Grosso for a number of inspiring and fruitful discussions related to the issues of this paper

7. Bibliographical References

Attia, M., Toral, A., Tounsi, L., Monachini, M., and van Genabith, J. (2010). An automatically built named entity lexicon for arabic. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C. (2006). Introducing the arabic wordnet project. In *Proceedings of the third international WordNet conference*, pages 295–300. Citeseer.

Cantineau, J. (1950). Racines et schémes. In *Mélanges offerts á William Marais par l'institut d'études islamiques de l'Université de Paris*. GP Maisonneuve, Paris.

Carter, M. (1990). Arabic lexicography. In J. D. Latham M. J. L. Youg et al., editors, *Learning and Science in the Abbasid Period ("The Cambridge History of Arabic Literature")*, pages 106–117. Cambridge University Press.

Del Grosso, A. M. and Nahli, O. (October 2014). Towards a flexible open-source software library for multi-layered scholarly textual studies: An Arabic case study dealing with semi-automatic language processing. In *Proceedings of 3rd IEEE International Colloquium, Information Science and Technology (CIST),Tetouan, Marocco*, pages 285–290, Washington, DC, USA. IEEE.

Dichy, J. (2002). Sens des schèmes et sens des racines en arabe : le principe de figement lexical (pfl) et ses effets sur le vocabulaire d'une langue sémitique. In L. Panier et al., editors, *La polysémie*. Presses Universitaires de Lyon.

Elkateb, S., Black, W., Rodríguez, H., Alkhalifa, M., Vossen, P., Pease, A., and Fellbaum, C. (2006). Building a wordnet for arabic. In *Proceedings of The fifth international conference on Language Resources and Evaluation (LREC 2006)*. Citeseer.

Fellbaum, C. (1998). *WordNet*. Wiley Online Library.

[4] The class u/u implies that the perfective scheme is R1aR2uR3a and the imperfective scheme is yaR1R2uR3u. In the other way, the second vowel is / u / in the both cases (the other vowels are the same to all verbs).

[5] Interlingual External Refs instead are contained within a so called Sense Axis component, and are used when resources contain senses or synsets from different languages (Francopoulo and George, 2008).

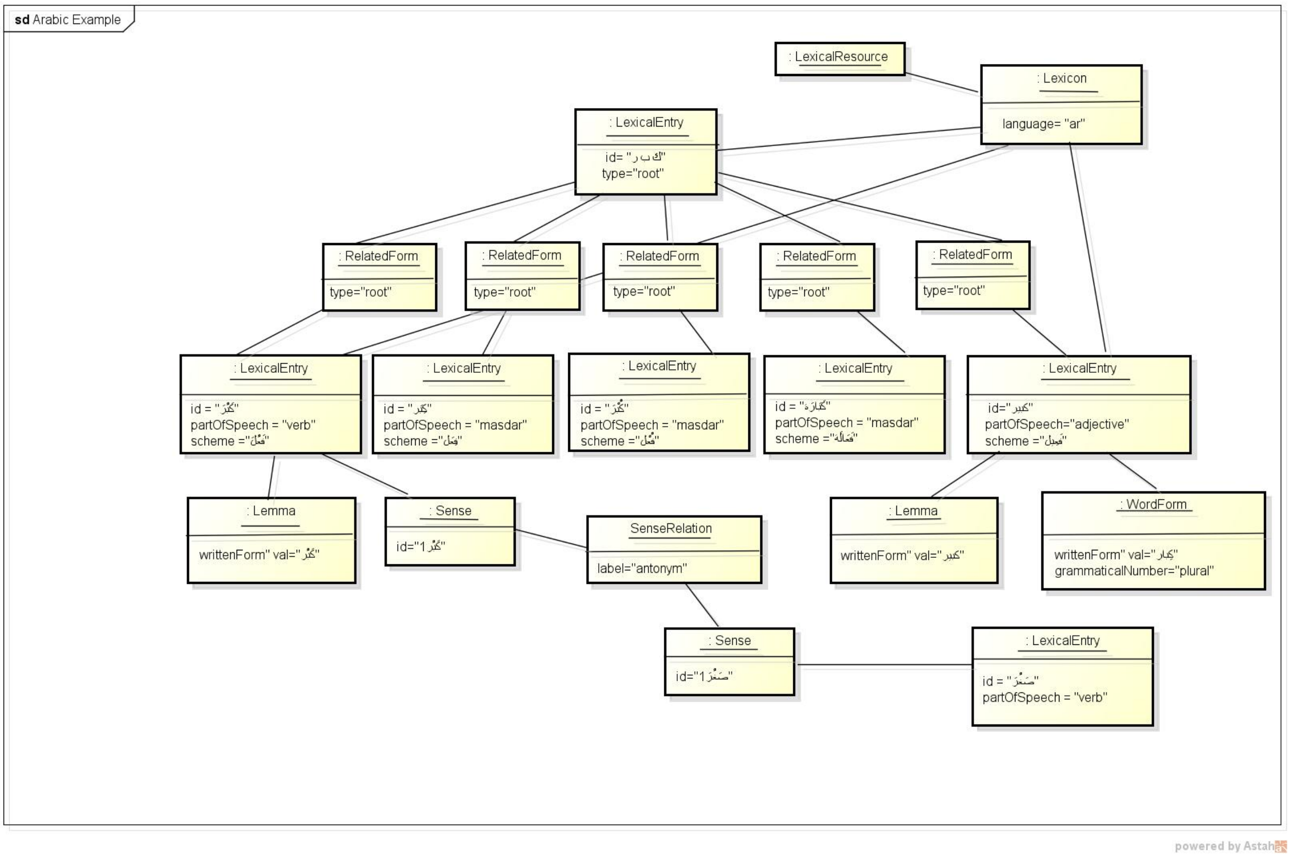


Figure 4: Partial schema of the LMF version of the lexicon.

Francopoulo, G. and George, M. (2008). Language Resource Management. *Lexical markup framework (LMF). Technical report, ISO/TC 37/SC 4 N453 (N330 Rev. 16*.

G. Francopoulo, editor. (2013). *LMF Lexical Markup Framework*. John Wiley & Sons.

Khemakhem, A., Gargouri, B., Abdelwahed, A., and Francopoulo, G. (2007). Modélisation des paradigmes de flexion des verbes arabes selon la norme LMF-ISO 24613 . In *Proceedings of TALN*.

Khemakhem, A., Elleuch, I., Gargouri, B., and Hamadou, A. B. (2009). Towards an automatic conversion approach of editorial Arabic dictionaries into LMF-ISO 24613 standardized model. In *Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt*.

Khemakhem, A., Gargouri, B., Haddar, K., and Ben-Hamadou, A. (2013). Lmf for arabic. In G. Francopoulo, editor, *LMF Lexical Markup Framework*, pages 83–98. John Wiley & Sons.

Kouloughli, D. E. (1991). *Grammaire de l'arabe d'aujourdhui*. Édition Pocket.

Lancioni, G. (1997). Sull'ordinamento dei dizionari arabi classici. *Rivista degli studi orientali*, 71:113–127.

Loukil, N. and Haddar, K. ahd BenHamadou, A. (2007). Normalisation de la représentation des lexiques syntaxiques arabes pour les formalismes dunification. In *Proceedings of the 26th conference on Lexis and Grammar*, Bonifacio.

A. Moutaouakil, editor. (1989). *Pragmatic Functions in a Functional Grammar of Arabic*. Foris Publications - Holland.

Pease, A., Niles, I., and Li, J. (2002). The suggested upper merged ontology: A large ontology for the semantic web and its applications. In *Working notes of the AAAI-2002 workshop on Ontologies and the Semantic Web*, volume 28.

Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M. A., Black, W., Elkateb, S., Kirk, J., Pease, A., Piek, V., and Fellbaum, C. (2008a). Arabic wordnet: Current state and future extensions. In *Proceedings of the Fourth International GlobalWordNet Conference - GWC 2008, Szeged, Hungary, January 22-25*.

Rodríguez, H., Farwell, D., Ferreres, J., Bertran, M., Alkhalifa, M., and Martí, M. A. (2008b). Arabic wordnet: Semi-automatic extensions using bayesian inference. In *Proceedings of the the 6th Conference on Language Resources and Evaluation (LREC 2008). Marrakech (Morocco)*.

Appendix

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "DTD_LMF_REV_16.dtd" >
<LexicalResource>

   <GlobalInformation>
      <feat att="info" val="The dictionary al-Qamūs is one of the most
       widely used in Arabic and it was compiled by Fairuzabadi, also known as
       El-Firuz Abadi or al-Fīrūzābādī (1329–1414)"/>
   </GlobalInformation>

   <Lexicon>
      <feat att="language" val= "ar"/>
      <LexicalEntry id= "كبر">
         <feat att="type" val="root"/>
         <Lemma></Lemma>
      </LexicalEntry>
      <LexicalEntry id = "كَبُرَ">
         <feat att="partOfSpeech" val="verb"/>
         <feat att="scheme" val="فَعُلَ"/>
         <feat att="inflectionalParadigm" val="u/u" />
         <Lemma>
            <feat att="writtenForm" val="كَبُرَ"/>
         </Lemma>
         <RelatedForm targets = "كبر">
            <feat att="type" val="root"/>
         </RelatedForm>
         <Sense id="كَبُرَ1">
            <SenseRelation targets="صَغُرَ_1">
               <feat att="label" val="antonym"/>
            </SenseRelation>
            <MonolingualExternalRef>
               <feat att="type" val="near_syn"/>
               <feat att="ExternalSystem" val="PWN3.1"/>
               <feat att="ExternalReference" val="200125649-v"/>
            </MonolingualExternalRef>
            <MonolingualExternalRef>
               <feat att="type" val="near_syn"/>
               <feat att="ExternalSystem" val="SUMO_ontology"/>
         <feat att="ExternalReference" val="FullyFormed"/>
      </MonolingualExternalRef>
   </Sense>
   <SyntacticBehaviour id="intransitive" />

</LexicalEntry>

<LexicalEntry id = "كِبَر">
   <feat att="partOfSpeech" val="masdar"/>
   <feat att="scheme" val="فِعَل"/>
   <Lemma>
      <feat att="writtenForm" val="كِبَر"/>
   </Lemma>
   <RelatedForm targets = "كبر">
      <feat att="type" val="root"/>
   </RelatedForm>
</LexicalEntry>

<LexicalEntry id = "كُبْرَ">
   <feat att="partOfSpeech" val="masdar"/>
   <feat att="scheme" val="فُعْل"/>
   <Lemma>
      <feat att="writtenForm" val="كُبْرَ"/>
   </Lemma>
   <RelatedForm targets = "كبر">
      <feat att="type" val="root"/>
   </RelatedForm>
</LexicalEntry>

<LexicalEntry id = "كَبَارَة">
   <feat att="partOfSpeech" val="masdar"/>
   <feat att="scheme" val="فَعَالَة"/>
   <Lemma>
      <feat att="writtenForm" val = "كَبَارَة"/>
   </Lemma>
   <RelatedForm targets = "كبر">
      <feat att="type" val="root"/>
   </RelatedForm>
</LexicalEntry>
```

```
    <LexicalEntry id="كَبِير">
      <feat att="partOfSpeech" val="adjective"/>
      <Lemma>
        <feat att="writtenForm" val="كَبِير"/>
        <feat att="scheme" val="فَعِيل"/>
      </Lemma>
      <WordForm>
        <feat att="writtenForm" val= "كِبار"/>
        <feat att="grammaticalNumber" val="plural"/>
      </WordForm>
      <WordForm>
        <feat att="writtenForm" val= "كُبَّارُون"/>
        <feat att="grammaticalNumber" val="plural"/>
      </WordForm>
      <RelatedForm targets = "كبر">
        <feat att="type" val="root"/>
      </RelatedForm>
    </LexicalEntry>
    <LexicalEntry id = "صَغُر">
      <feat att="partOfSpeech" val="verb"/>
      <feat att="scheme" val="فَعُلَ"/>
      <Lemma>
        <feat att="writtenForm" val="صَغُر"/>
      </Lemma>
      <Sense id="1_صَغُرَ"/>
    </LexicalEntry>
  </Lexicon>
</LexicalResource>
```