



**HAL**  
open science

## ALIF editor for generating Arabic normalized lexicons

Samia Ben Ismail, Hajer Maraoui, Kais Haddar, Laurent Romary

► **To cite this version:**

Samia Ben Ismail, Hajer Maraoui, Kais Haddar, Laurent Romary. ALIF editor for generating Arabic normalized lexicons . The International Conference on Information and Communication Systems, IEEE Jordan Section, Jordan University of Science and Technology, Apr 2017, Irbid, Jordan. hal-01469966

**HAL Id: hal-01469966**

**<https://hal.science/hal-01469966>**

Submitted on 16 Feb 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ALIF editor for generating Arabic normalized lexicons

Samia Ben Ismail  
ISITCom Hammam Sousse  
MIRACL Laboratory  
Sfax, Tunisia  
[Samia\\_benismail@yahoo.fr](mailto:Samia_benismail@yahoo.fr)

Hajer Maraoui  
Faculty of Science of Sfax  
MIRACL Laboratory  
Sfax, Tunisia  
[hajer.maraoui@gmail.com](mailto:hajer.maraoui@gmail.com)

Kais Haddar  
Faculty of Science of Sfax  
MIRACL Laboratory  
Sfax, Tunisia  
[kais.haddar@yahoo.fr](mailto:kais.haddar@yahoo.fr)

Laurent Romary  
Inria, Team Alpage  
[Laurent.romary@inria.fr](mailto:Laurent.romary@inria.fr)

**Abstract**—the development of a normalized morpho-syntactic Arabic lexicon is not an easy task. In fact, many norms allow the structuration and representation of lexical data. The adoption of a stable standard will guarantee the interoperability and interchangeability of lexical resources. Still, research work that deals with normalization for Arabic lexical resources is not well developed yet, especially for some standards such as the TEI (Text Encoding Initiative). In this context, we aim at creating an Arabic lexicon editor with a constraint checker based on both the ISO standard LMF (Lexical Markup Framework) and the TEI guidelines. To develop this editor, we use a linguistic approach composed of several steps. The editor’s prototype named ALIF can guarantee the construction of two types of output lexicon files: one in LMF and the other in TEI. The evaluation of this system is based upon a lexical database that contains all the derived and inflected forms generated from a lexicon of 10 000 canonical verbs. The results obtained were encouraging despite some flaws related to exceptional cases of difficult words.

**Keywords**— LMF; TEI; Lexicon Editor; Interoperable formats

## I. INTRODUCTION

The standardization of lexical information becomes an obvious necessity, especially for linguistic communities to cover the compatibility and the interchangeability between NLP applications. In fact, the Arabic-speaking community that deals with lexica adapts this idea to create standardized lexical databases favouring the integration, the evaluation and the interoperability of lexical resources. However, creating such lexical databases is not an easy task for a linguist user. In fact, a linguist has no computing development knowledge to create standardized lexical resources based on existing norms. Moreover, even if he/she is familiar with the standards structures, the creation of lexica can only be processed manually and this task takes a long time. Therefore, the existence of a lexicon editor based on standards models and formats can simplify the description of the lexical databases and comply with certain constraints especially when it endowed with a constraints checker. Besides, the lexical editor can be the basis for computer based training activities and support non-expert users. The realisation of such an editor raises a number of serious problems. The first major difficulty was encountered in the preparatory phase. In fact, it relates with the selection and identification of data categories from standards registers used by the processing

levels (i.e., morphological, syntactical, etc.) and the language specification. The second difficulty occurs when looking for a generic approach to generate a variety of types of normalized lexicons such as LMF (Lexical Mark-up Framework) and TEI (Text Encoding Initiative) simultaneously. Another serious problem lies in the identification of lexical constraints (for example inflectional ambiguity may appear in several conjugation rules for verbs with the same scheme).

Our main objective is to create an LMF/TEI editor dedicated to the generation of Arabic lexica covered from the morphological, syntactic and semantic levels. To realise this objective, we made an in-depth study of Arabic words (verbs, nouns and particles) on two levels: lexical and syntactic.

Besides, to reach lexical standardization, we based our work on the ISO standard 24613 LMF and the TEI guidelines (<http://www.tei-c.org/>). We have worked with LMF for the following reasons: First, LMF allows the specification of monolingual and multilingual lexica. In addition, it insures a stretchable and modular modelling covering all the levels of linguistic description (morphological, syntactic and semantic). Furthermore, its flexibility of modelling favours the representation of the characteristics of Arabic language in particular the derivational and inflectional aspects. We also choose to work with TEI because it has clear, simple and concrete structure. Further, the encoding of textual and lexical content with TEI is based upon a well-maintained schema and offers a large spectrum of possible elements. Moreover, it is easy to identify a subset of the TEI which is compatible with LMF [1].

In the present paper, we begin with an overview of the LMF and TEI standards. These two standards are necessary for the encoding of the Arabic lexica for Modern Standard Arabic (MSA). Then, we present the specificities of the MSA in the perspective of the modelling of a lexicon. After that, we present the experiment of the approach by a prototype and its evaluation.

## II. OVERVIEW ON LMF

ISO 24613 LMF [2] is a standard whose scope is the standardization and the representation of monolingual, bilingual and multilingual lexical databases. The LMF standard allows the management of the data exchange between these resources. It also allows the merging of a large number of lexical resources.

LMF conforms to the specifications and main modelling of UML defined by OMG. In fact, it is a meta-model represented by UML classes and the associations between them.

According to ISO 42613, LMF is decorated by data categories selected from a register (DCR). It consists of a core and eight extensions. Meta-model core (see Figure 1) is a lexical entry backbone. It specifies the notions of lexicon, word, form and sense. It admits extensions describing the various descriptions levels of the monolingual and multilingual lexical resources. All these extensions cannot be represented independently of the core.

In the NLP domain, several works have focussed on the lexicon standardization. Among the works dealing with Arabic lexica on the basis of the LMF standard, we can mention for example the development of ArabicLDB [3]. The ArabicLDB aims at the construction of a conjugation system of the Arabic verbs and nouns. The problem of this system is that it is not endowed with a constraints checker and it does not handle all the descriptions levels such as syntactic level.

We can also point to an approach for projecting one or more syntactic lexicons existing HPSG to LMF [4]. This prototype can produce either lexica corresponding to LMF without any loss of information, or lexica corresponding to LMF with loss of information and also can produce lexicons not corresponding to LMF with loss of information. Also, there exists another approach for projecting LMF lexicon of Arabic verbs to TDL lexicon [5]. This prototype can add as much transformation rules in the rule file as necessary, without changing the code. But the major complexity lies in the number of rules used and the size of the XML file given as input.

After our study of the existing systems, we find that these systems are complicated to use and also are not free. Besides, they do not take the specificities of Arabic language such as LEXUS [6] and Marphalou [7]. Furthermore, they work only with LMF norm. However, these inconveniences did not prevent that these works are the source of inspiration to realise our prototype.

### III. OVERVIEW ON TEI

The TEI is an international consortium whose goals are the development of a set of standards for the preparation and the exchange of electronic texts. This initiative was established in 1987 by a set of researchers working on archiving, structuring and of analysis of the electronic texts [8].

The publication of the works of the various committees' research was in the form of "Guidelines" [9]. Indeed, the TEI recommendations present conventions of usable coding in several domains. They can be applied as well to create new information that to exchange existing information. A TEI document has to comply with XML coding rules. The

fundamental structure of a TEI document bases itself on a structure which can be, according to the project BVH, a printed text or another a manuscript [10]. This structure describes the textual part of the document. The notion of structure is essential to edit exchange and analyze texts thanks to the TEI guidelines. In what follows, we give overviews onto the structure and documents in TEI representations.

Several works are made throughout the process of development and improvement of the TEI from its first recommendations until the last publications. However, these works have not dealt strictly the lexica. Besides, there is an absence of Arabic language treatment although the TEI has been the basis for many projects on documents with various natural languages. One of the TEI projects works on dictionary structures. Dictionaries appear among the complex document types the most handled by the TEI. The workgroup of the TEI on dictionaries had occupied the production of a set of conventions at the level of dictionary entries. Every entry of a dictionary is a strongly structured object, in which the numerous mechanisms of abbreviation and typographic organization allow a presentation condensed by the information [11] [1].

Dictionary entries contain several clearly recognizable types of information: information on the form of the word (spelling, pronunciation, caesura, etc.), grammatical information (word class, morphology, etc.), definitions or translations, etc. To normalize lexica with TEI, we started from the idea of considering lexica as a dictionary with a high level of details where each entry can get morphological, syntactical and semantic descriptions. We inspired from the TEI dictionary structure to customize a lexicon structure dedicate for Arabic lexicons [12].

### IV. PROPOSED METHOD FOR LEXICAL ENCODING

In this section, we explain the various operations for the word transformation in the inflected forms. These operations can be an addition of a letter, a deletion or a replacement of consonant or brief vowel. The employment of these operations defers according to schema and category [13]. Fig. 1 specifies the operations for the conjugation of a *sound hamzated* [صحیح [مهموز] verb with the *singular third person* [هو] (he) in the *inaccomplished* aspect with the *passive* voice.

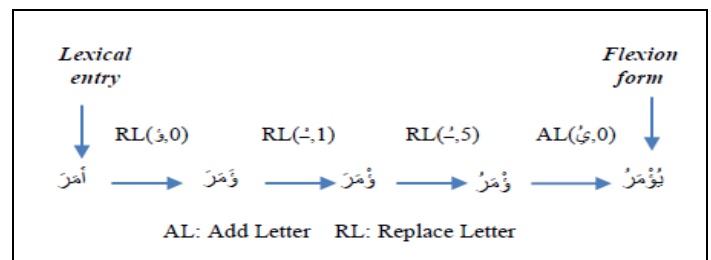


Fig. 1. Generation of inflectional form (يُؤَمَّرُ) /y?maru/.

Fig. 1 explains the operations to generate an inflected form for the *sound hamzated verb* (أمر) (to order) of schema [فَعَلَ-يَفْعُلُ].

The extraction of the inflected forms for nouns (as the dual and plural forms) is made by the same operations of the inflection paradigms for verbs [14].

Fig. 2 illustrates the essential operations to produce the *broken plural* of the *singular noun* (قَلَمٌ) (pencil).

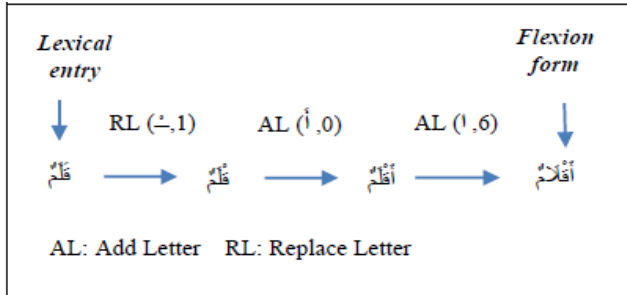


Fig. 2. Calculation of inflected form (أَقَلَمٌ) /ʔqlamun/.

Inflection paradigms are the necessary operations to generate the inflected / derived forms and to introduce the conjugation paradigms [15]. To reach the determination of the inflected/derived forms for a verb, it is necessary to pass by a process of operations. The appearance of these operations depends on the category of the handled word. The structure of the rule files used by this prototype postponed from a standard to another one. Files appropriate to LMF standard are built by an XML model however files appropriate to TEI standard are built by TEI model.

## V. IMPLEMENTATION

To implement our prototype, we used some tools. First, we designed our system with UML. Then, we used Oxygen XML Editor to create the rule files and to propose a TEI structure to describe Arabic lexicons. After that, we developed the editor interfaces using JAVA language and the API JDOM Library.

### A. Lexicon creation

The module of lexicon creation consists of two sub-modules: one is to describe generally the lexicon and the other one is for adding lexical entries. In the following part, we are going to explain the mechanism of this module.

The first stage of creation of new lexicon is the choice of the lexicon entry type (verbs, nouns or particles) and the lexicon output file type (TEI, LMF or both of them). In the second stage, the user is going to select the properties of his lexicon. These properties are the authentication properties (workspace and password) and lexicon properties (lexicon name and

corresponding package). The second sub-module is for the adding of the lexical entry description in the lexicon file. In this part, ALIF editor connects to the rules and constraints files to generate a morphological description (derived and inflected forms) for each entry given by the user and according to its properties (scheme and category). Besides, ALIF editor offers the choice for the user to create lexicon by adding the canonical entries one by one or all of them in an external text file. For the example of a verb lexicon, Fig. 3 illustrates the interface to create new verb lexicon.



Fig. 3. New verb lexicon interface.

To generate derived and inflected forms for each verb, ALIF editor selects the specific rules matches with the verb properties from the rule files. Then, the system applies the derivational and inflectional rules one by one on the canonical form of the verb. ALIF editor generates a list of TEI/LMF output files for the created lexicon and save them in the user workspace. The number of the output files depends on the number of canonical forms given by the user. ALIF editor starts with creating an output LMF/TEI file containing the first morphological description for all the canonical forms (category of the word, schema, root, pronunciation, etc.). For each canonical form, the editor creates another LMF/TEI file containing all the derived and inflected forms. Every canonical form gets an attribute to mention the reference of the correspondent inflectional and derivational file. For example, for 500 verbs the ALIF editor generates 1002 output files for each lexicon type. All these files are saved in one package in the user workspace.

### B. Lexicon management

In this module, the user can update his lexicons. This module includes an authentication step. After choosing the lexicon and entry type, the system passes to the authentication step to verify the properties of the user. After that, the system launches an interface named "Open Lexicon" which permits three services. The first one allows consulting the created lexicon. The second permits to look for a lexical entry in the chosen lexicon. The third allows adapting the selected lexicon. The first service

allows the consultation of the selected lexicon file. That will be by opening one of the created TEI or LMF lexicon files. Fig. 4 presents a fragment from TEI lexicon for the derived and inflected forms of the verb [عَلَّمَ] / ʕʕallama / (to teach).

```

<entry>
<form type="lemma">
  <orth type="vocal">عَلَّمَ</orth>
  <orth type="unvocal">علم</orth>
  <pron>ʕʕallama</pron>
  <gramGrp>
  <pos>Verb</pos>
  <case>فعل ثلاثي مزيد و هو صحيح سالم</case>
  <subc>فعل ناقص متعدي إلى مفعولين</subc>
  <gram type="scheme">فَعَّل</gram>
  </gramGrp>
</form>
<form type="inflected">
  <orth type="vocal">تَعَلَّمَ</orth>
  <orth type="unvocal">تعليم</orth>
  <pron>taʕʕlii.mun</pron>
  <gramGrp>
  <pos>مصدر</pos>
  <gram type="scheme">تَفَعَّل</gram>
  </gramGrp>
</form>
<form>
  <orth type="vocal">عَلَّمْتُ</orth>
  <orth type="unvocal">علمت</orth>
  <pron>ʕʕallamtu</pron>
  <gramGrp>
  <per>1</per>
  <number>sing</number>
  <gram type="tense">الماضي</gram>
  <gram type="grammaticalVoice">المجهول</gram>
  </gramGrp>
  <form type="pronom_personel">
  <orth>أنا</orth>
  </form>
</form>
<form>
  <orth type="vocal">عُلِّمْتُ</orth>
  <orth type="unvocal">علمت</orth>
  <pron>ʕʕullimtu</pron>
  <gramGrp>
  <per>1</per>
  <number>sing</number>
  <gram type="tense">الماضي</gram>
  <gram type="grammaticalVoice">المجهول</gram>
  </gramGrp>
  <form type="pronom_personel">
  <orth>أنا</orth>
  </form>
</form>
</entry>

```

Fig. 4. Fragment from the TEI verb lexicon.

Fig. 5 presents a fragment from LMF lexicon for morphological and syntactic level of the verb [عَلَّمَ] / ʕʕallama/ (to teach).

```

<LexicalEntry name="عَلَّمَ" morphologiquePatterns="فعل تام
متعدي بمفعولين" id="علم">
<Stem>
<feat att="writtenForm" val="عَلَّمَ"/>
<formRepresentation>
<feat att="writtenForm" val="ʕʕallama" />
<feat att="writtenForm" val="latn" />
</formRepresentation> ... </Stem>
<syntacticBehaviour>
<subcategorizationFrame>
<feat att="description" val="فعل+اسم+مفعولين بدون أداة"/>
</syntacticBehaviour>
<feat att="partOfSpeech" val="verb" />
<feat att="root" val="ع ل م" />
<feat att="scheme" val="فَعَّل" />....
<Relatedform>
<feat att="writtenForm" val="مُعَلَّم" />
<feat att="catGram" val="اسم مفعول" />
<feat att="sheme" val="مُعَلَّل" />
...
</Relatedform> ...
</LexicalEntry>
<morphologicalPatterns target="فعل تام متعدي بمفعولين"
id="علم">
<feat att="patternType" val="flexionnalPattern" />
<feat att="example" val="عَلَّمَ" />
<feat att="partOfSpeech" val="verb" />
...
<TransformSet val="عَلَّمْتُ">
<feat att="comment" val="تصريف الفعل مع أنا في الماضي"/>
<grammaticalFeatures val="أنا">
<feat att="grammaticalTense" val="ماضي"/>
<feat att="grammaticalGender" val="neutral"/>
<feat att="person" val="1"/>
<feat att="grammaticalNumber" val="singular"/>
<feat att="grammaticalVoice" val="معلوم" />
</grammaticalFeatures>
</TransformSet> ...
</morphologicalPatterns>

```

Fig. 5. Fragment from the LMF verb lexicon.

Fig. 6 presents a fragment from the LMF noun lexicon of the noun [وَلَدٌ] /waladun/ (a boy).

```

<LexicalEntry name="وَلَدٌ" id="ولد">
...
<feat att="partOfSpeech" val="noun" />
<feat att="scheme" val="فَعَّل" />
<Lemma>
  <feat att="type" val="إسم ثلاثي مجرد و هو صحيح سالم" />
...
</Lemma>
</LexicalEntry>

```

Fig. 6. Fragment from the LMF noun lexicon of the noun [وَلَدٌ] /waladun/ (a boy).

```

<morphologicalPatterns id="ولد">
  <feat att="patternType" val="flexionnalPattern" />
  <feat att="example" val="وَلَدٌ" />
  <feat att="partOfSpeech" val="اسم مشترك" />
  ...
  <TransformSet val="أولاد">
    <feat att="comment" val="جمع تكسير" />
    <grammaticalFeatures>
      <feat att="grammaticalGender" val="مذكر" />
      <feat att="grammaticalNumber" val="جمع" />
    </grammaticalFeatures>
  </TransformSet>
</morphologicalPatterns>

```

Fig. 6. Fragment from the LMF noun lexicon.

From the lexicons generated by the editor, the user has the option of exposing the inflected forms of the lexicon in table form or in normalized XML file with TEI or LMF like the examples of the figures 5 and 6. The second step allows the user to look for a lexical entry in the selected lexicon. The editor extracts the selected entry descriptions from the lexicon and shows them to the user.

In this section, we present the main interfaces of ALIF editor. These interfaces are shared between the creation of a lexicon or lexical entries and the management of the existing lexicons. In the following part, we are going to evaluate our ALIF system.

## VI. EVALUATION AND DISCUSSION

To evaluate our editor prototype, we created and tested a normalized syntactical lexicon containing 10000 Arabic verbs and 500 nouns. Consequently, for the verb lexicon, the prototype generates respectively 20002 TEI and LMF files representing the 10000 canonical forms. From these canonical forms ALIF editor's system generates 1 130 000 inflected and derived forms (113 forms per verb). Table 1 illustrates the obtained results. For the nouns lexicon, ALIF creates 1002 TEI and LMF files representing 4500 inflected forms for the 500 nouns (9 forms per noun).

TABLE I. SUMMARY TABLE OF THE SYSTEM RESULTS.

Lexicon type	Obtained forms	Correct forms	Erroneous forms
Verb lexicon	1 130 000	1 129 700	300
Noun lexicon	4500	4350	150

Table 1 shows that sometimes for irregular verbs, we can obtain erroneous generated forms. For the verbs lexicon, the total number of obtained forms is 1 130 000. The program succeeded to produce 1 129 700 correct forms. However, we found 300 erroneous forms requiring rectification. This problem of erroneous forms is related with some irregular verbs. The

examples (1) and (2) illustrate a case of irregularity of verb conjugation.

(1) خَذَ → أَخَذَ      (2) أَوْجِرَ → أُوجِرَ  
 /ʔa/xa/δa/ → /xu/δ/      /ʔa/dʒ/ra/ → /ʔu/ʔ/dʒ/r/

The verb [أَخَذَ] (to take) how has the schema [فَعَلَ-يَفْعُلُ] is an irregular verb in the imperative with the personal pronoun [أَنْتَ] (you).

For the noun lexicon, the total number of the obtained forms is 4500. The editor succeeded to create 4350 correct forms and 150 erroneous. Failure cases are (in most of the cases) due to the non-detection of the inflectional rules for plural nouns of some nouns that are irregular in inflection. For example, the nouns (كُتَّابٌ - شَاعِرٌ - شَعْرَاءٌ) are in the same category and have the same scheme but they did not have same inflectional rule for plural.

Indeed, we estimate the quality of our work manually. Table 2 illustrates the obtained values of precision, recall and F-score. According to the value of precision, we conclude that the value of precision for the verb lexicon is worth 0.99. For the noun lexicon, the precision is equal to 0.96. Also, the recall value for the verb lexicon is 0.99 and for the noun lexicon 0.96. These values provide an F-measure equal to 0.49 for verb lexicon and 0.48 for noun lexicon.

TABLE II. SUMMARY TABLE OF THE PRECISION, RECALL AND F-SCORE.

	Precision	Recall	F-score
Verb Lexicon	0,99	0,99	0,99
Noun Lexicon	0,96	0,96	0,96

Consequently, we conclude that the obtained results are encouraging. Besides, we can say that this prototype is flexible and easy to maintain because it is based on an object-oriented programming language. However, we handled some problems. Some of them are related with the irregular verbs and nouns which need to integrate heuristics solving this kind of problems.

Other kind of problems was related with the optimization of response time. We experience the creation of a verb lexicon by adding entries one by one. Then we found that the execution time augmented with the number of entries and it takes much time to create a lexicon with more than 500 entries. So, we tried to run the system with other method and start with a text file contains the entries. This solution makes the execution faster and much easier for the user.

## VII. CONCLUSION

In this paper, we have elaborated a lexicon editor supporting two standards: LMF and TEI. We have also integrated a module

for constraint checker. In this editor, we have proposed a type hierarchy for Arabic words according to a done study on Arabic word specificities. As mentioned, the elaborated editor allowed us to generate two normalized lexicons with large coverage for Arabic words: one for LMF and the other for TEI. This editor can be considered also as a good tool for linguists dealing with normalization. The obtained values of measures show that the results obtained from our editor are encouraging. These results can be used in others levels of analyses. Moreover, the execution time is correct. Indeed, the lexicon files need only few minutes to be generated.

As perspectives, we want to improve the proposed models. This improvement consists on incorporating new elements describing more deeply the various levels of description. Besides, we want to develop more our TEI model to incorporate a syntactical description and to get to a deep morphological level. Moreover, we want elaborate a tool allowing the transformation from LMF lexicons to TEI ones and vice versa. To increase the visibility of our lexicon editor, we want to post it open source.

#### REFERENCES

- [1] L. Romary, "TEI and LMF crosswalks," In INRIA & HUB-IDSL, hal-00762664, version 1, 2015.
- [2] G. Francopoulo, "LMF Lexical Markup Framework," ISTE LTD. Wiley New York, 1st ed, 2013.
- [3] A. Khemakhem, B.Gargouri, A.Ben Hamadou and G.Francopoulo, "ISO Standard Modeling of a large Arabic Dictionary," Journal of Natural Language Engineering, Cambridge University Press, 2015.
- [4] K. Haddar, H.Fehri and R.Laurent, "A prototype for projecting HPSG syntactic lexica towards LMF," In Journal of Language Technology and Computational Linguistics Gesellschaft für Sprachtechnologie und Computerlinguistik.V, (GSCL), n°27 (1), pp. 21–46, 2012.
- [5] N. Loukil, K. Haddar and A. Benhmadou, "A Syntactic Lexicon for Arabic Verbs," in proceedings of Seventh International Conference on Language Resources and Evaluation LREC, Malta, pp.269-272, 2010.
- [6] M. Kemps-Snijders, M.-J.Nederhof and P.Wittenburg, "LEXUS, a web-based tool for manipulating lexical resources," In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006), pp. 1862-1865, 2006.
- [7] L. Romary, S. Salmon-Alt and G. Francopoulo, "Standards going concrete: from LMF to Morphalou," COLING, Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries, the 20th International Conference on Computational Linguistics, Switzerland, pp. 22-28, 2004.
- [8] L. Burnard and C.M . Sperberg-McQee, "TEI P5: Guidelines for Electronic Text Encoding and Interchange," Text Encoding Initiative Consortium, Version 3.0.0. revision 89ba24e, 2016.
- [9] N. Ide and J.Veronis, "Une application de la TEI aux industries de la langue: le Corpus Encoding Standard," In Cahiers GUT enberg, n°24, pp. 166–169, 1996.
- [10] N. Dufournau, M-L. Demonet and T. Uetani, "Manuel d'encodage XML-TEI Renaissance et temps modernes Imprimés-manuscrits," Version Beta, UMR 6576, 2008.
- [11] L. Romary and W. Wegstein, "Consistent modeling of heterogeneous lexical structures," In Journal of Text Encoding Initiative, Issue 3, TEI and linguistics, 2012.
- [12] H. Maraoui and K. Haddar, "Automatisation de l'encodage des lexiques arabes en TEI," In 2nd conference on CEC-TAL, Sousse, Tunisia, 2015,
- [13] A. Dahdeh, "المعجم الوسيط في تصريف الأفعال," Lebanon Librerie, Lebanon, 1999.
- [14] A. Dahdeh, "معجم قواعد اللغة العربية في جداول و لوحات," LibrerieNachirun. Lebanon, 5th ed, 1997.
- [15] S. Ammar and Y.Dichi, "الشامل في تصريف الأفعال العربية," Bescherelle collection. Hatier, Paris, ISSN 0990 3771, 1999.