Design of a Synthesizer and a Semantic Analyzer's Multi Arabic, for use in Computer Assisted Teaching

Khairedine Bacha and Mounir Zrigui

Laboratoire UTIC, Équipe de Monastir Faculté des sciences de Monastir
E-mail: Khairi.bacha@gmail.com, Mounir.Zrigui@fsm.rnu.tn

Abstract

In this paper, we describe a multifunctional electronic dictionary which contain a number of applications that cover the most of linguistic levels (morphology, syntax, semantics.), size and complexity satisfactory, based on the Arabic language, generated and labeled automatically and contains an amount of lexical information at all levels and all aspects of the behavior of words, respecting the proper organization and simple to facilitate its operations.

Keywords: Electronic Dictionary, Language Levels, Arabic

Introduction

Production and development of electronic dictionaries of morphemes-roots have become the privileged object of man-machine communication. Indeed, as these activities are the research areas of the first order in artificial intelligence and computational linguistics. The language industries are emerging, however, studies conducted in recent years in various countries, particularly in France and Canada, show that these industries are expected to grow dramatically.

If electronic dictionaries have undeniable advantages over paper dictionaries (maneuverability, reduced size), they are far from perfect in terms of content and information access. Coverage (number of entries) is surely not the only criterion of a good dictionary, because what good is to have a rich database if it is not possible to easily and data quickly?

The work we undertake within the scope of the automatic processing of Arabic, it aims to automatically generate dictionaries of Arabic.

The need for Natural Language Processing (NLP) in large lexical resources continues to intensify. Management of such knowledge must be considered a priority because it is fundamental to the success (in the sense efficiency) of NLP applications
that use them. Also, she leads generally, much of the cost of these applications. Thus, increases the interest of development of lexical databases reusable and independent of a particular language implementation.

On the other hand traditional electronic dictionaries allow nowadays, thanks to advanced queries, quick access to articles and information they contain. However, it disappears once the user succeeded in the article because he finds the presentation of the paper versions and their limits, the other on the Arabic language is because it is a language difficult to automatically process, For this there are a number of prototype or experimental systems for the Arabic language.

Firstly, we show here, through our example, a new dictionary for learning the Arabic language (as complete as possible), a derivative, a conjugator and a morphological analyzer for Arabic words [1], [2] and [3], how it is possible to organize differently dictionary articles, more dynamic and more responsive to user needs. This is made possible by a flexible and coherent structuring of lexical data. Why we decided to perform an electronic dictionary size and complexities satisfactory, based on the Arabic language, and labeled automatically generated and contains a large amount of lexical information at all levels and any aspect of behavior of words in respecting the proper organization and simple to ease its operation. In a second step, we used these tools to create a number of educational applications for Arabic

**Science Dictionaries**
The dictionary (Science dictionaries) which is a branch of natural language processing, has been one of the interesting activities for Arab linguists, but in their time, they treated paper dictionaries. They began by grouping terms according to their areas as the books ""elasmai"" "كتب الأسماعي في الخليل وال_mix"" and then they designed the dictionary using a method that facilitates identification of words and their meaning. It is believed that ""الخليل بن أحمد الفراهيدي ""which is the first Arabic language but a dictionary that called it ""آتاب العين "". Several linguists have appeared after that and took several dictionaries, dictionaries of these we quote

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Author</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>(الحروف) مجموع</td>
<td>لابي عمرو الشبياني</td>
<td>ت 206 هـ</td>
</tr>
<tr>
<td>(الألفاظ) مجموع</td>
<td>لابن السكيت</td>
<td>ت 244 هـ</td>
</tr>
<tr>
<td>(الجملة) لاين دريد</td>
<td>ت 321 هـ</td>
<td></td>
</tr>
<tr>
<td>(البر) طريقي</td>
<td>لابن علي القاني</td>
<td>ت 356 هـ</td>
</tr>
<tr>
<td>(المجهر) نظري أهداف</td>
<td>لابن فارس</td>
<td>ت 395 هـ</td>
</tr>
<tr>
<td>(المصطلحات) لابن سعد</td>
<td>لابن الريحي</td>
<td>ت 400 هـ</td>
</tr>
<tr>
<td>(المصطلحات البحري)</td>
<td>لابن سعد</td>
<td>ت 458 هـ</td>
</tr>
<tr>
<td>(المصطلحات المحدود)</td>
<td>لابن سعد</td>
<td>ت 711 هـ</td>
</tr>
<tr>
<td>(المصطلحات المحدود)</td>
<td>لابن سعد</td>
<td>ت 817 هـ</td>
</tr>
<tr>
<td>(المصطلحات المحدود)</td>
<td>لابن سعد</td>
<td>ت 1205 هـ</td>
</tr>
</tbody>
</table>
Electronic Dictionaries
According to Bogaards [4], the standard dictionary does not help the learner understand the text and to know its meaning, since the learner hate the use of the dictionary because it distracts from the process of reading text, and more it does not attempt to exert efforts to find the word in question in the dictionary, then dice the existence of a small indication, he accepts it. Finally, the dictionary has a negative effect on the reading task, and according to the experiments, the dictionary adds no great change.

Dice 1990, the public could use different electronic dictionary [5], the use of dictionary is easier than paper dictionaries because of improvements in lexical access and how to navigate in the dictionary

- The Lexical Access: For a foreign learner of the language, the use of dictionary is a bit difficult because the selected text, the word processing is an inflected form of a root, hence the transformation of this inflected form to canonical form, which represents the dictionary entry, is not easy, at all, for a special language such as Arabic, which is characterized by the difficulty of handling a morphological analyzer. So why, he found two possible solutions, or to list the possible canonical forms of inflected form in question, which account for the electronic dictionary entries, or use a morphological analyzer well understood, that little extract the root represents the entry in the dictionary.

- Interactivity: Helps to navigate the electronic dictionary in order to improve the efficiency to find the word you look for. The electronic dictionary has to be improved compared to the standard dictionary, point of view of the display method, which is connected to user selection to display such information and such, this reduces the problem that is already in the standard dictionary, and an easier route of entry.

Nowadays, it is not enough to adapt to electronic dictionaries old, bringing to the text of a dimension for which he had not been designed. The electronic dictionary is actually a creation in itself and requires the adaptation of work lexicographic requirements and possibilities of computer technology: the combination of sound and image to the textual definition, the dimension offered by analog navigation in hypertext, the ease and speed of consultation..

On the other hand, the a priori unlimited storage of information in electronic form allows you to add dimensions to the dictionary which previously gave rise to different structures: grammar explanations, declensions, conjugations, spelling, analog references, synonyms, homonyms can be retrieved from a single screen for consulting a dictionary language. Historical developments, soundtracks, video, photographs, paintings summary three-dimensional images can enrich an electronic encyclopedia.

Construction "specialized" dictionaries for machine Translation
The ATLAS software [6] [7] designed by Daniel Bachut allows to introduce new words and their associated codes in a machine translation dictionary. He manages
indexing manuals for linguists. His code was written in Pascal and it was compiled on
a VM / ESA IBM. Dictionaries filled with ATLAS are used by the translation system
Ariane-G5.

The manual indexing are decision trees, used by lexicographers at the indexing of
an entry. They explain how to assign different codes used in the translation. The
linguist publishes its manual with any text editor. It then compiles it with ATLAS.
Once ATLAS detects an error, it reports its position and allows the linguist to correct
it. Once the compilation is finished, the linguist can use functions to add cards in its
manual or delete interactively. It can also view and print all or part of its manual. The
manual can also be visualized as a tree.

Statistical study
In what follows, we present a statistical study on several dictionaries da Arabic (المحيط
االمحيطن, محطة المحيطنات ) we presented information on the
number of words, their occurrences in the dictionary (عدد نماد,متوسط عدد المواد للحرف
1) (عدد الكلمات بالمعجم, المشتقات

<table>
<thead>
<tr>
<th></th>
<th>عدد الكلمات بالأصل</th>
<th>عدد المشتقات بالأصل</th>
<th>متوسط عدد المواد بالأصل</th>
<th>المتحف</th>
</tr>
</thead>
<tbody>
<tr>
<td>المحيط</td>
<td>810.000</td>
<td>40.000</td>
<td>40.000</td>
<td>1.429</td>
</tr>
<tr>
<td>الغني</td>
<td>1.300.000</td>
<td>84.965</td>
<td>11.200</td>
<td>400</td>
</tr>
<tr>
<td>المحيط</td>
<td>450.000</td>
<td>30.000</td>
<td>7000</td>
<td>250</td>
</tr>
<tr>
<td>الغني</td>
<td>2.000.000</td>
<td>195.000</td>
<td>30.000</td>
<td>1.071</td>
</tr>
<tr>
<td>القاموس</td>
<td>733.000</td>
<td>70.000</td>
<td>11.000</td>
<td>390</td>
</tr>
<tr>
<td>المحيط</td>
<td>4.493.934</td>
<td>158.149</td>
<td>9.393</td>
<td>335</td>
</tr>
<tr>
<td>نجعه القدر</td>
<td>119.176</td>
<td>5.629</td>
<td>142</td>
<td>---</td>
</tr>
</tbody>
</table>

Some Research Projects on Dictionaries
Computerized Dictionary of Arabic and Multilingual Corpus Based on
(DIINAR-MBC)
DIINAR (computerized dictionary of Arabic) [8] [9] [10] [11] [12] is one of the most
important lexical databases in the field of NLP treating Arabic. It is characterized by a
good level of coverage, about 20,000 entries and 39,000 verbal nominal inputs.
DIINAR designers organize information of various kinds in different databases
depending on the nature of the token processed. DIINAR.1 [13] [14] used primarily to

---

1 www.ajeeb.com
generate a lexicon for use by applications of NLP as it is an intentional representation that is based on rules. On the other hand, lexicons generated from DIINAR have content and different shapes. The build environment can meet different needs without affecting the structure of the lexical database.

**DicoBase™ Arabe**
The Arab Linga DicoBase™ is a dictionary-type database for Arabic language used in applications of computational linguistics. The Arab DicoBase™ is the result of years of work in the field of terminology and lexicography. The team is composed of several Linga Doctors in Arabic literature and terminologistes professionals under the direction of Professor Leila-Guillemot.

**Different Methodologies**
There are different methodologies for creating and maintaining dictionaries complex [15]. We can mention the construction:
- **Direct**: using the interface of a database,
- **Democratic**: using an editor of the trade as "pseudo-syntax editor" ,
- **Classic**: publishers using SGML syntax type,
- **Specialized**: using a real ad hoc syntax editor created for the dictionary in question,
- **Online**: by contributors working directly on the Web

**Table 3: Different methodologies**

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Directe</th>
<th>Democratic</th>
<th>Classical</th>
<th>Specialized</th>
<th>Line</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Benefits</strong></td>
<td>• First method&lt;br&gt;• Using a commercial database</td>
<td>• Simplicity&lt;br&gt;• Distribution of work&lt;br&gt;• Word software evolution</td>
<td>• Widely used method&lt;br&gt;• First method of constructing dictionaries for human use</td>
<td>• Comfort for the lexicographer&lt;br&gt;• Structure computer simplified</td>
<td>• Simultaneous work in groupware</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>• Blocking of data&lt;br&gt;• No inconsistency in the database</td>
<td>• No tools for consistency checking&lt;br&gt;• Need for administrator lexicologist</td>
<td>• No direct onlinemust equip all editors of an appropriate editor.</td>
<td>• Difficulty in maintaining an ad hoc tool&lt;br&gt;• Complexity</td>
<td>• Valid only for dictionaries of simple structure&lt;br&gt;• Need for tools to the editor</td>
</tr>
<tr>
<td><strong>Example</strong></td>
<td>• ((French-English)-Malaysian 4D</td>
<td>• Word RTF</td>
<td>• Canadian bilingual dictionary</td>
<td>• ATLAS&lt;br&gt;• DECID</td>
<td>• Project SAIKAM</td>
</tr>
</tbody>
</table>
Realization of a Prototype Multifunction Dictionary

Method of construction of dictionary

The construction of a dictionary can be divided into several tasks as speakers:

In practice, there can be a lexicographer who is at the same time lexicographer. In Subsequently, however, we will use the term lexicographer to describe the person who defines the information contained in the dictionary but also who controls the work of lexicographers (editor of the publication of a dictionary).

Creating tools for dictionaries problematic computer mostly from the mass and variety of information to build. The construction of a dictionary is a collaborative work by different lexicographers who must meet a consistency not just in regard to the form specified by the lexicographer (abbreviations, tags ...), but also on the bottom (same criterion selection effect, the same criteria and inputs decomposition subentries in the case of homo-graphs ...).

Finally, the choices made by some lexicographers can influence decisions to be taken by other lexicographers (syntactic or semantic links between entries). The tools built should therefore take into account the distributed aspect of the work of lexicography.

During work of lexicography, it may happen that the lexicographer wants to change the structure of the dictionary to better take into account some phenomena that have been mispriced or undervalued. This can result in a change of editing interfaces and a possible modification of automatic verification tools for consistency. A tool for lexicographers must be sufficiently scalable and configurable to authorize such changes.

Under the TELA project (Towards Environmental Learning Arabic)) [16], we designed and implemented a prototype dictionary. This prototype consists of about
8000 lexical units (verbs, nouns and adjectives). He resumed, morphological and syntactic information previously developed. Based on this prototype, we developed an electronic dictionary for learning that we operate within a learning environment [16] We first made tools for automatic processing of Arabic who will attend human expert in the choice of vocabulary and definitional examples of textual corpus. Then we tried to address issues raised regarding the use of different types of dictionaries.

Architecture
This architecture is derived from two decompositions

- A functional decomposition which is to decompose the system into subsystems, trying to regroup in the same set of functions consistently.
- A static decomposition for describing each subsystem by describing the various modules of it. It results in a subgraph and also defines the interaction of the subsystem with the other. Each module thus obtained by this decomposition can be divided in turn into basic units.
- On the lexicographic, our principles are based on a large Arabic dictionary, "Mohit DICTIONARY EL" which are the basis of our differentiator conjugator Standard Arabic.

![General architecture of semantic analyzer multifunction](image)

**Figure 2:** General architecture of semantic analyzer multifunction

Choice of Method
The role of the specification stage is to express what must be the software and not how it should do, why we have chosen the SADT method because it allows not only to describe the project tasks and their interactions, but also to describe the system that
the project will explore, create or modify, in including highlighting the parts that make up the system, purpose, operation of each, and the interfaces between the various parts that make a system is not a mere collection of independent elements, but a structured organization of these in a specific purpose.

La méthode du hash-code généralisée permet le meilleur temps d’accès, et compte tenu de la grande quantité d’informations dans le dictionnaire, nous avons choisis la méthode du hash-code généralisé, en mettant l’accent sur la minimisation du temps d’accès

**Calculation and increase the indicative**

We take \( I = \sum_{i=1}^{n} \text{Ord}\ (r_i) \cdot B^{i-1} \)

\( r = r_1 \ldots r_n \) (n root length)

\( B: \) basis for a better dispersion of the roots

Are
- \( IF = \min(\text{ord}\ (r_i)) \leq i \leq n \)
- \( IM = \max(\text{ord}\ (r_i)) \leq i \leq n \)

We have:

\[
\sum_{I=1}^{n} IF \cdot B^{i-1} \leq \sum_{I=1}^{n} IM \cdot B^{i-1}
\]

\( \Rightarrow IF \cdot [\sum_{I=1}^{n} B^{i-1}] \leq IM \cdot [\sum_{I=1}^{n} B^{i-1}] \)

\( \Rightarrow IF \cdot (1-B^n)/(1-B) \leq IM \cdot (1-B^n)/(1-B) \quad B > 1 \)

**Visual overview of the Application**

We developed a module to allow the learner to view information acquired in a multi-user dictionary. Our work has shown the importance of this module to support the learner to record and organize the vocabulary in part known. It is therefore entirely appropriate to consider this information to the lexical specification of activities and choice of text corpus.

To prevent the user dictionary is a reproduction of the general dictionary, we added three new fields that help define the synonym and an illustration of the lexical unit acquired a first part allows the learner to take lessons general courses, the second is to diversions of Arabic verbs, the third deals with verb conjugations, apart the old business of an electronic dictionary.

In Arabic, the majority of verbs and nouns are combinations of a root (usually with three radicals) and a pattern. A word family can be generated automatically on the same semantic concept from a single root using different reasons, by the way, even Arab roots can be generated automatically.
With this electronic dictionary, we can easily determine the roots of words, then these roots can be used to help learners of Arabic to find synonyms and word meanings. Because in order to find a definition or synonym from an Arabic word, and without human assistance, learners of Arabic will use the thesaurus of the paper (in alphabetical order of the roots of differentiable words) to deal with texts incomprehensible.. Pour identifier un mot dans le dictionnaire il faut connaître ses racines, ce qui n'est pas évident pour une personne qui apprend l'arabe.

*Access Techniques to the Dictionary*

If electronic dictionaries have undeniable advantages over paper dictionaries (maneuverability, reduced size), they are far from perfect in terms of content and information access. Coverage (number of entries) is surely not the only criterion of a good dictionary, because what good is a rich database if it does not allow easy and fast access to data?

*To do this, we developed two modules*

The module analyzer, which analyzes the morphosyntactic word entered by the user. The integrated module 'researcher' who will find the dictionary filtering roots obtained after analysis of the word introduced and find corresponding information. This module is integrated inside the module analyzer.'
The module 'ANALYZER'

This module requires several processing blocks. These blocks are both independent and linked to each other. The following diagram shows the different 'blocks' of treatment which shows the path of the analyzer:

![Diagram of processing blocks](image)

**Figure 4:** The different processing blocks 'analyzer'

The first function of our system is the morpho-syntactic analysis of the word introduced. For this we made a module analyzer ', which will contribute to access and
make it easier and faster retrieval of information. The analyzer processes a word any introduced by the user. Its main aim is to isolate the root word, since in general to find a word in Arabic electronic dictionary, you must know the root and then go in a certain order, the list of entries starting with the first character of this root.

![Diagram of a dictionary search interface](image)

**Figure 5:** Search a word in the dictionary

Our system uses this module to access dictionary information. In our case, we have linked lists and tables, containing the entries and related information (the inflected words or derivatives, their properties and characteristics. Access to these memory structures is done either for checking the validity of a word (derived from a root determined) entered by a user, or for information concerning any one derivative (input word) of the root.

The integrated module 'SCIENTIST' inside 'the ANALYZER'

Once the root word introduced found the ANALYSER 'module uses the' researcher '. This module allows you to filter the different roots candidates, if not directly seek information corresponding to the root, if the system has found one.
Racinisation using prefix and suffix
This method works racinisation by using an algorithm that makes the separation of prefixes and/or suffixes that characterize a pseudo-root without taking account of infixes, the token is found the root form of a pattern of Arabic language [17], this approach is called the racinisation light. All algorithms performed to arrive this approach have the same defect as they are impassable. This approach is based on the determination of prefixes and suffixes using a predetermined list, see table below for removed or if they are at the beginning of a word which looks for the root.
Table 4: List of prefixes and suffixes [18]

<table>
<thead>
<tr>
<th>Prefixes</th>
<th>Suffixes</th>
</tr>
</thead>
</table>

The general principle of operation of the module 'researcher' is as follows:
- From the triliteral root obtained, the researcher points to the list, with entries beginning with the first letter of the root.
- Browse (found) entries, until the corresponding input (ie find the same root). Une fois que la racine Once the root sought is found, the 'researcher' points to all its corresponding derivatives.
- Can the 'researcher' points bending codes corresponding to each of the derivatives found. These codes describing the bending derivatives (where the derivative is either a verb, a noun or adjective).
- Transmit the results (information) to our system.

The different processing blocks of the module 'CONJUGATION' Method used

The principles of suffixation and prefixation of words enabling the conjugation of verbs, as shown in the diagram below. So we would be here too tempted to automate the operation and completely automatically generate inflected forms.
This procedure first generation system consists of two steps

- First step: manual
- Second step: automatic
Design of a Synthesizer and a Semantic Analyzer's Multi Arabic

It aims to build all the verbal inflected forms of the Arabic language. It is to conjugate all the verbs in all aspects and with all the pronouns. Here in more detail the method we have advocated.

**Determination of all forms of conjugation**

This is to have a table that contains all the codes with their triliteral verbs conjugation. We know that in the Arabic language, there are several forms of verb conjugation naked. Conjugate a verb, knowing its root and form number, is simply to select the structure that corresponds to conjugation, and will replace the letters" wildcards" by radical letters of the verb, so it must be determined for each verb exist in our database the exact form of the combination of his own.

To number the forms of triliteral verb conjugation, were consulted "انتشار التصريف" [19], which allowed us to determine the different forms of combination codes.

**Example**

<table>
<thead>
<tr>
<th>Normal</th>
<th>Deaf</th>
<th>Hamzée 1</th>
<th>Hamzée 2</th>
<th>Hamzée 3</th>
<th>Concave Waw</th>
<th>Concave Ya</th>
<th>Defective Waw</th>
<th>Defective Ya</th>
<th>Assimilated Waw</th>
<th>Assimilated Ya</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Deaf</td>
<td>Hamzée 1</td>
<td>Hamzée 2</td>
<td>Hamzée 3</td>
<td>Concave Waw</td>
<td>Concave Ya</td>
<td>Defective Waw</td>
<td>Defective Ya</td>
<td>Assimilated Waw</td>
<td>Assimilated Ya</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>38</td>
<td>41</td>
<td>43</td>
<td>123</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>98</td>
<td>117</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>38</td>
<td>42</td>
<td>43</td>
<td>123</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>99</td>
<td>118</td>
</tr>
</tbody>
</table>

**Figure 8:** Example forms of conjugation with their codes

Arabic Verbs is relatively simple. It is performed identically to the bare verbs and derived verbs. In Arabic, we speak of aspects of modality and voice [20].
Two aspects: it is the done (الماضي) expressing that the action is completed, and the Imperfect (المضارع) expressing that the action is not completed.

Three modes: they exist only for the Imperfect.

It was indicative (ا), the subjunctive (منصوب) and apocopated (مجزوم). The other side has only one method: the area code.

Other modes of the Arabic verb conjugation, but we opted to keep only the major modes.

Wo votes for the accomplished and unaccomplished: the active voice (المعلوم) and the passive voice (المجهول).

Another peculiarity of the Arabic verb, is that in addition to singular and plural pronouns, it is combined with pronouns duels giving rise to the 13 pronouns.

This procedure generation system is completely automatic. It aims to build all the Arabic verbal inflections. It is to combine all verbs in all aspects and with all the pronouns.

**Bypass module**

Another feature that seems important for learning Arabic is the derivation. Indeed it is the most important mechanism in the generation of Arabic lexicon. The relationship between morphology and lexicon are the words that must be learned with all the information to their reuse. The act of transforming verbal and nominal derivatives predicts three grammatical behavior, namely the accusative, the nominative and genitive and associate these states to determine.

![Processing module 'BYPASS'](image)

**Figure 9:** Processing module 'BYPASS'
The starting point of our generation chain verbal roots of the dictionary is certified that we have already generated, the chain will result an electronic dictionary containing verbal inflections. The generation procedure is expertly automatic, while the testing phase and filtration remains manual. One of the requirements that we wanted to give our generation chain is to minimize manual operations for the benefit of automatic procedures, so the original proceeding in this chain has been designed to allow the entry of a data volume which is minimal, but so far that allows to generate automatically and reliably a dictionary as comprehensive as possible. At the beginning of Arabic morphological system, there are the verbal roots, if we except a few cases where there is ambiguity, it is enough to have patterns of verbs and the root from which it is to automatically determine the verbs in question. Also decree that instead of us enter the verbs themselves, it will designate their schemes. We therefore propose the operator field derivational "virgin" in the form of a list of schemes and will be asked to activate the patterns that go with the root introduced.

<table>
<thead>
<tr>
<th>The root and derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>حسبَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
<tr>
<td>عَطَالَة</td>
</tr>
</tbody>
</table>

**Figure 8:** Derivational field of a root

**Module 'TRANSLATION**

The source language of our machine translation system and the various components of a system of statistical machine translation. In this section, we will implement and test
a machine translation system that manipulates data enriched with morphosyntactic information for the Arabic / English, Arabic / French and propose some original solutions to improve the referral system.

Recall that the approach to statistical machine translation is as follows. Given an Arabic sentence $s$, we seek the English translation of $c$ that maximizes $p(s \mid c)$, the probability that a sentence is the translation of $s$ (we always translate Arabic into English $s$ in this $c$ following).

The figure 9 represents the main components of probabilistic machine translation system. The decoder takes as input the source text, the translation model and the language model for outputting the translated text. Note that the language into which we want to translate is called "target language".

To be able to test and validate the overall approach, the first version of machine translation module was performed. It allows the creation of virtually any type of words and phrases, without collection and use of the trace of their implementation by the learners. The screen shot below shows the current state of the "visible" module developed so far. The ergonomics and design interfaces are provisional and should evolve in the future.

**Figure 9**: Machine statistical machine translation Arabic / English
Examples of Activities
We designed twenty activities that continuously changes with each call, eliminating the cases dealt with later. Most modules contain support activities that are available after a certain time given to the learner. This assistance takes the form of indices depends on the activity such as word length and the first letters, so as to guide the learner towards the correct solution.

The following example (see figure below) represents an analysis of words, the application chosen at random a word of Arabic, using the random function, from the dictionary. The learner must choose a model for this word in a list box, enter the root and see if his answer is correct or not. This interface is used to assess learner knowledge.

Figure 10: Arabic Translation – English
The result reflects the activities of partially correct answers with the diagnosis according to adapt the answers of the learner. In fact, an answer is not always completely false and can be decomposed into several elements, some are proven correct.

Results of Activities
Traces recording the shares on which the learner can make a synthesis of system usage. The traces are organized by session work and record any action taken by the learner date to calculate the time spent in each response in all activities and a saw to the teacher about how the system is used and assess whether such an activity must be modified.

In what follows an example of the use of these results

Figure 11: Example of activity
Table 5: Results of activity

<table>
<thead>
<tr>
<th>Spent time</th>
<th>Learner 1</th>
<th>Learner 2</th>
<th>Learner 3</th>
<th>Learner 4</th>
<th>Learner 5</th>
<th>Learner 6</th>
<th>Learner 7</th>
<th>Learner 8</th>
<th>Learner 9</th>
<th>Middle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response1</td>
<td>1</td>
<td>0.9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.933333333</td>
</tr>
<tr>
<td>Response2</td>
<td>1</td>
<td>0.8</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.911111111</td>
</tr>
<tr>
<td>Response3</td>
<td>1</td>
<td>0.5</td>
<td>0.9</td>
<td>0.8</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>1</td>
<td>0.877777778</td>
</tr>
<tr>
<td>Response4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Response5</td>
<td>1</td>
<td>0.3</td>
<td>0.6</td>
<td>0.5</td>
<td>1</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>0.722222222</td>
</tr>
<tr>
<td>Response6</td>
<td>1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.8</td>
<td>0.588888889</td>
</tr>
<tr>
<td>Response7</td>
<td>0.8</td>
<td>0.5</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>0.6</td>
<td>0.777777778</td>
</tr>
<tr>
<td>Response8</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>0.888888889</td>
</tr>
<tr>
<td>Response9</td>
<td>0.6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.944444444</td>
</tr>
<tr>
<td>Response10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Activity</td>
<td>9.4</td>
<td>7.3</td>
<td>7.7</td>
<td>8.5</td>
<td>10</td>
<td>8.5</td>
<td>8.5</td>
<td>9</td>
<td>8.9</td>
<td>8.644444444</td>
</tr>
</tbody>
</table>

The evaluation of an activity can be accomplished using several parameters (number of correct responses over time to respond, the appreciation of the learner, ...). In our example, the evaluation is done by time spent per learner per response.

Figure 12: Graphical result of the activity

This table allows the teacher to conduct a comparative study at the level of understanding of each learner. So for our example, we note that 2 is the best learning and learner-5 is the lowest in this activity. This graph also shows that this activity is not reached, because we can see that the success rate lower responses to recent issues,
which generally opposes the objective of the activity (more success in the late activity) and shows that the educational and pedagogical purpose of this activity is not reached.

**Conclusion and outlook**

The aim of our work was to design and implement multi-semantic analyzer for Arabic, based on NLP tools with unlimited vocabulary goals and use the simplest possible. This will work, then the starting point for new opportunities for improvement and refinement on the basis of comparative tests and evaluations. To this end, we were led to make choices, to meet a number of problems.

The next step is to ensure that the system can make a diagnosis "evolved" typos or demanding answers detection approach. The system should also provide some assistance to facilitate inference and to refer the student to the correct solution. Therefore, these exercises will be more then a few simple tests, but an educational, activity that promotes learning of Arabic vocabulary.

**References**


