ABSTRACT

Irregular verbs can be defined as verbs that act differently from the basic patterns in all or some cases; those verbs may act irregularly in the derived forms as well, irregular verbs pose a challenge to Machine Translation (MT) in particular; morphological and syntactical analysis, the definition of irregular verbs involves accounting doubled, hamzated and weak verbs. This paper is presenting work-in-progress to build rule based machine translation with lexical properties and characteristics to handle Arabic irregular verbs agreement in English-Arabic MT. Arabic lexicon would be supported with a strong theoretical framework and implemented using robust tools that will facilitate its implementation. in this paper we built a module to detect irregular verbs, i.e, doubled, hamzated, mithal, hollow, defective, and enfolding. A set of rules have been conducted based on the tense of the verb, place of the vowel root-letter, (i.e., first, second or third person representation), number, gender, tense and mood features, Our proposed module has been effectively evaluated using real test data and achieved satisfactory results.

Keywords: MT, agreement, irregular-verbs, hamzated, doubled, hollow, defective, word-ordering, perfective, imperfective.

1. INTRODUCTION

Arabic verbs are constructed on a root that uses three consonants or radical latters which is known as (Morphological balance) (C₁aC₂aC₃a)¹. This non-concatenative way used in this paper is typically called root and pattern morphology - to identify the basic meaning of the verb. However, the inflection of verbs in Arabic is mainly achieved through the use of prefixes and suffixes denoting person, number, gender, case, mood and tense.

The major two categories of verbal structures in Arabic are the perfective and the imperfective. The imperfective is used for both the simple present and simple future tenses to denote an unfinished act which is still in progress, while the perfective is used for the simple past tense, however, the perfective verbs indicate a completed act.

Arabic imperfective verbs have five moods as shown in figure 1. four of which (the indicative, subjunctive, jussive, and energetic) share one structure but with different endings. The fifth mood, the imperative, has its own distinct structure.

There are Seven morpho-syntactic features involved in the agreement of Arabic irregular verbs: number (singular, dual and plural), gender (feminine and masculine), person (1st 2nd, and 3rd), case (nominative, accusative and genitive), definiteness (definite and indefinite), mood (indicative, imperative, subjunctive, jussive and energetic) and tense (perfect, imperfect and participle) as shown in figure 1. [12].

According to (Wightwick, J and Gaafar M., 2008) Weak verbs are the largest category of irregular verbs. They can be subdivided into four types depending on which of the root letters is affected:

- Verbs with waaw or yaa’ as the first root letter (Mithal verbs).
- Verbs with waaw or yaa’ as the second root letter (Hollow verbs).
- Verbs with waaw or yaa’ as the third root letter (Defective verbs).
- verbs that have two weak letters in their roots (Enfolding verbs).


² To clarify the structure of Morphological forms we have used the corresponding CV array of each form alongside. Cᵣ corresponds to radical letters, and represent the consonants of فعل.
2. REVIEW OF LITERATURE

Arabic is the fourth most widely spoken language in the world. It is a highly inflectional language, with a rich morphology, relatively free word order, and two types of sentences [2][3].

Verbs in Arabic are categorized in different ways according to the needs of the grammarians or applications. Verbs can be classified depending on the number of characters that form their root, or depending on the nature of characters forming their root as this will influence their conjugation and the forms of their derivations [6].

Verbs can be further sub-categorized by tense (past, present and future), case (nominative, accusative and genitive), with respect to transitivity (intransitive and transitive), aspect (perfective, imperfective and imperative), with respect to the subject (person, number and gender) and, voice (active and passive).

Corbett (2001) defined agreement as “systematic covariance between a semantic or formal property of one element and a formal property of another.”, he used the terms “controller” to refer to the element which determines the agreement, “target” to refer to the element whose form is determined by agreement, and “domain” to refer to the syntactic environment in which agreement occurs [15][24][31].

Attia (2008) stated that Arabic has rich agreement morphology which allows it to show agreement relations between various elements in the sentence.

Essentially, the Arabic word can be described as follows: [prefix1][prefix2] stem [infix] [suffix1] [suffix2] [1], table 1 below shows some examples of the affixes handling.

<table>
<thead>
<tr>
<th>Suffixes</th>
<th>Suffixes</th>
<th>Stem</th>
<th>Prefixes</th>
<th>Prefixes</th>
<th>Arabic</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>C1aC2aC3ia</td>
<td>C1aC2aC3ia</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>yaC1C2C3ia</td>
<td>yaC1C2C3ia</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>yuC1C2C3ia</td>
<td>yuC1C2C3ia</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>yaC1C2C3iham</td>
<td>yaC1C2C3iham</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>yaC1C2C3iham</td>
<td>yaC1C2C3iham</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>syaC1C2C3iham</td>
<td>syaC1C2C3iham</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>syuC1C2C3iham</td>
<td>syuC1C2C3iham</td>
</tr>
<tr>
<td>s1</td>
<td>s2</td>
<td></td>
<td>s1</td>
<td>s2</td>
<td>syuC1C2C3iham</td>
<td>syuC1C2C3iham</td>
</tr>
</tbody>
</table>

Suffixes in Arabic can be categorized into two basic forms, the suffixes that are attached to the verbs and the suffixes that are added to the nouns [32]. Furthermore, some of the suffixes can be attached to both the noun and verb stem. Nevertheless, Arabic permits the use of up to three suffixes simultaneously to be attached to the end of the same stem [1]. Furthermore, Arabic words are built from roots rather than stems and involve diacritization. Written Arabic
is also characterized by the inconsistent and irregular use of punctuation marks [16]. Table 2 below presents a wide range of suffixes example for the verb hit (ضرب).

**Table 2.** Arabic suffixes examples (adopted from Abu Shquier. M and Abu Shqeer. O 2012) [6]

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Suffix description</th>
<th>Suffix category</th>
<th>Example</th>
<th>phonetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>نع</td>
<td>First person</td>
<td>Verb</td>
<td>ضربني</td>
<td>drbny</td>
</tr>
<tr>
<td>ك</td>
<td>Second</td>
<td>Both</td>
<td>ضريك</td>
<td>drbk</td>
</tr>
<tr>
<td>ه</td>
<td>Third person</td>
<td>Both</td>
<td>ضريه</td>
<td>drbh</td>
</tr>
<tr>
<td>حا</td>
<td>Third person</td>
<td>Both</td>
<td>ضريها</td>
<td>drbha</td>
</tr>
<tr>
<td>هم</td>
<td>Third person</td>
<td>Both</td>
<td>ضريهم</td>
<td>drbhm</td>
</tr>
<tr>
<td>كم</td>
<td>Second</td>
<td>Both</td>
<td>ضريكم</td>
<td>dbkm</td>
</tr>
<tr>
<td>كن</td>
<td>Second</td>
<td>Both</td>
<td>ضريبك</td>
<td>dbkn</td>
</tr>
<tr>
<td>كما</td>
<td>Second</td>
<td>Both</td>
<td>ضريهما</td>
<td>drbma</td>
</tr>
</tbody>
</table>

**3. LINGUISTICS**

Arabic language plays a crucial role with the root (C₁aC₂aC₃a) to add subtle variations to the meaning. There are nine significant derived forms categorized into four groups (for the singular masculine 3rd person in the present tense) as shown in table 3 below:

**Table 3.** Nine significant derived forms

<table>
<thead>
<tr>
<th>Group #</th>
<th>Form #</th>
<th>Structure</th>
<th>Arabic imperfect</th>
<th>Arabic perfect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>Form I</td>
<td>yaC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td>Group II</td>
<td>Form II</td>
<td>yuC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td></td>
<td>Form III</td>
<td>yuC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td></td>
<td>Form IV</td>
<td>yuC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td>Group III</td>
<td>Form V</td>
<td>ytaC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td></td>
<td>Form VI</td>
<td>ytaC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td>Group IV</td>
<td>Form VII</td>
<td>yanC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td></td>
<td>Form VIII</td>
<td>yaC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
<tr>
<td></td>
<td>Form X</td>
<td>ystaC₁aC₂aC₃a</td>
<td>فعل</td>
<td>فعل</td>
</tr>
</tbody>
</table>

- Forms II and IV can have the meaning of carrying out an action to someone/something else.
- Forms II and IV are making the verb transitive or causative.
- Form II can also give a verb the meaning of doing something intensively and/or repeatedly.
- Form III often carries the meaning of “doing something with someone else”; or the meaning of “trying to do something”.
- Form V is often the reflexive of form II.
- Form VII is relatively uncommon and usually has a passive meaning.
- Form VIII is a common form and is usually intransitive, these verbs are often close to the basic root meaning.
- Form X often has the meaning of asking for something.

**4. DISCUSSION AND APPROACH**

The proposed approach will examine the basic verbal forms of Arabic and determine which parts are determined by what. A fully inflected form of an Arabic verb may consist of prefixes, a stem and suffixes/postfixes. The suffixes are person and number agreement features while the prefixes depend on conjunctions. The stem is consisted of arrangements of consonants and vowels that indicate the root, the tense and the mood.

The following is an explanation of the model processes with an example:

**Process 1:** Receives the source text (English statement), and pass it to the parser;

(The girls saw the paper).

**Process 2:** Identifies POS by consulting the English grammar database table:

(The/DT girls/NNS saw/VBD the/DT paper/NN).

**Process 3:** Retrieves Arabic meanings as well as subject features from the English lexicon database table;

(The/DT girls/NN نبات/NN رأى/NN the/DT paper/NN).

**Process 4:** detect irregular verbs based on their categories.

**Process 5:** analyze subject-verb conjugations and apply rules to handle the derivation and inflection of Arabic irregular verbs (process is carried out with the help of a bi-lingual dictionary)

**Process 6:** Analyzes the source text semantically to decide whether SVO should be used or not;

(The result will be either yes or no).

**Process 7:** Creates the correct derivation of the equivalent Arabic regular verbs depending on the results

![Figure 2. Morphological Analysis with process 5](image-url)
from processes 3 and 4, and the consultation of the Arabic grammar, Arabic lexicon.

**Process 8:** Creates the correct derivation of the equivalent Arabic irregular verbs depending on the result of the 30 cases shown in the figure below and from processes 3 and 4, and the consultation of the Arabic grammar, Arabic lexicon.

(IF the result of process 4 is no, then the verb will be رأيت (C1aC2at) since the default mode VSO will be used; if the result is yes, the verb will be رأينَ (C1aC2aC3na) since SVO mode will be used).

**Process 9:** Finally, the complete Arabic translation is produces by referencing the words ordering rules database table:

The Arabic generator will synthesis the inflected Arabic word-form based on the morphological features to produce the surface Arabic target language (based on the result of the previous processes, we will get either رأيت البنات الىرقة “ra?at albnat alwaraqah” in the case of using VSO or البنات رأينَ الىرقة “albnat ra?aina alwaraqah” in the other case SVO).

5. RESULTS AND DISCUSSION

An experiment is conducted on 171 independent test suites based on the following aspects of agreement and ordering of irregular verbs:
1. Article-Noun Agreement.
2. Adjective-Noun Agreement.
3. Verb-Subject Agreement.
4. Demonstrative-Noun Agreement.
5. Relative Pronoun-Antecedent Agreement.
6. Predicate-Subject Agreement.
7. Order of the adjective.
8. Successive words form an expression.
10. Conjunction with "and"
11. Adding the preposition "ال".

After we classified the problems that cause that ill-agreement and ordering of Arabic irregular verbs we then compare between the output of three available systems and compare the output with the original translation of the input text based on those twelve problems. The result of the experiment is shown in the following table.

<table>
<thead>
<tr>
<th>No.</th>
<th>ALKAF</th>
<th>GOOGLE</th>
<th>TARJIM</th>
<th>Our System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage</td>
<td>92.1%</td>
<td>84.6%</td>
<td>94.2</td>
<td>96.1%</td>
</tr>
</tbody>
</table>

Some of the output of the three systems investigated here is found to give full coherent meaning, and in some cases grammatically correct, translations in general have a problem in one or more aspects. Incoherent translations seem to be produced due to some fault or deficiency in one (or more) phase or component of the analysis of irregular verbs.

5. CONCLUSIONS

This paper has dealt with irregular verb derivation in English-Arabic Machine Translation in conjugation with sentence word order. Through this paper we have explored the characteristics of Arabic language that will affect the development of a Machine Translation (MT). Several distinguishing features of Arabic pertinent to irregular verbs have been explored in detail with reference to some potential difficulties that they might present.

Our proposed module has been effectively evaluated using real test data and achieved satisfactory results. We concluded that we can enhance the output quality of English-Arabic MT by feeding the system with adequate, robust and completed rules to deal with the morph-syntactic inflectional morphological features of irregular verbs. To achieve this task we proposed a set of 30 rules based on the tense of the verb, place of the vowel root letter, first, second or third person representation, number and gender features, and diacritics preceding vowel letter, i.e., nominative, accusative or genitive case.

Through the investigation of the available MTs and related researches, as well as the flexibility of Arabic language grammars, we concluded that we are a bit far away from getting an English-Arabic MT up to the accuracy of human translation due to either faulty analysis of the SL text or faulty generation of the TL text.

6. REFERENCES


Al-Najem, S., Inheritance-based Approach to Arabic Verbal Root-and-Pattern Morphology


Feras Saeed. 2010 Word Order and Agreement Asymmetry (English and Foreign Languages University, Hyderabad). The 5th annual Moscow Student conference on Linguistics


