

Arabic irregular verbs: inflection and derivation for English-Arabic Machine Translation

Mohammed Mahmoud Abu Shquier
Department of IT, University of Tabuk
Tabuk, 41791, KSA

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 letters which is known as (Morphological balance) (C₁aC₂aC₃a)^{1 2} فعل - 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).

¹ The reference that we have used as sources of data is related to the Arabic Morphological Balance: Al Rajihi (1993), Wright (1967), Al-Hamalawi (1991), Omer et. al. (1984), Yaqub(1988), Makram (1987), Al-Dahdah (1991), Al-duqur (1986), and Mustafa et. al. (1989).

² To clarify the structure of Morphological forms we have used the corresponding CV array of each form alongside. C_{ns} corresponds to radical letters, and represent the consonants of فعل

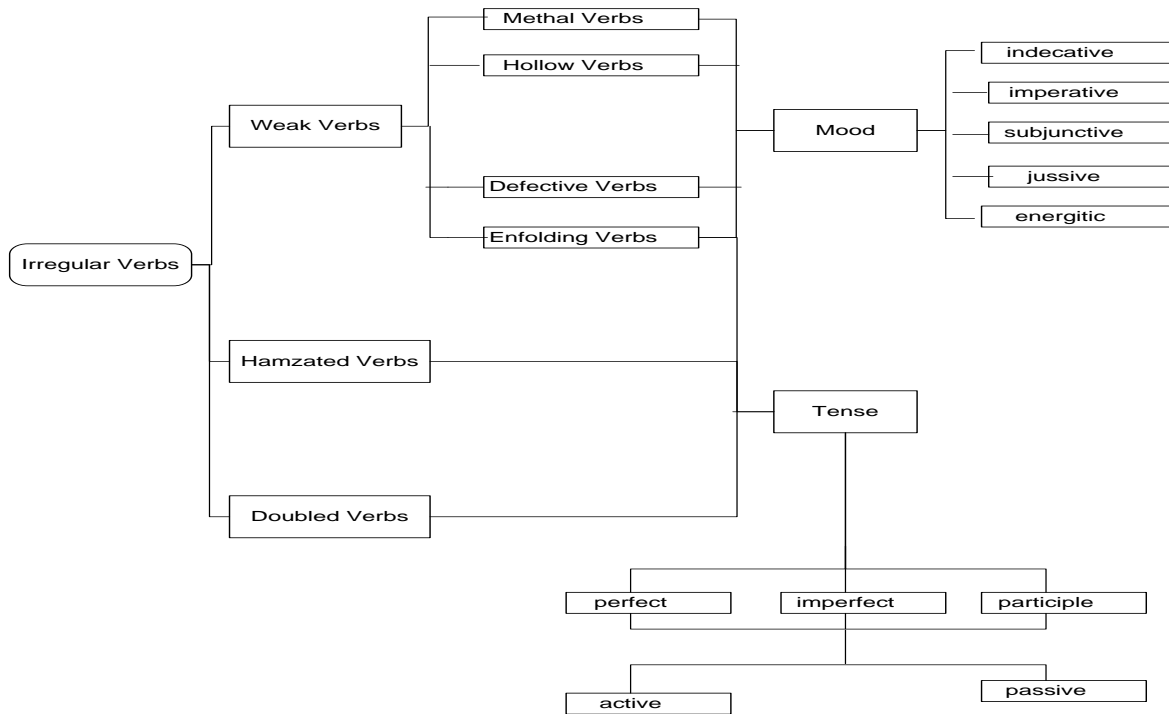


Figure 1. Arabic irregular verbs with their conjugation dependences

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][prefix1] stem [infixes] [suffix1] [suffix2] [1], table 1 below shows some examples of the affixes handling.

Table 1. an Arabic affixes/suffixes examples.

suffixe s2	suffixe s1	infix es	stem	Prefixes 2	prefixe s1	Arabic word	Structure
-	-	-	ضرب	-	-	ضرب	C ₁ aC ₂ aC ₃ a
-	-	-	ضرب	ي	-	يضرِب	yaC ₁ C ₂ iC ₃ u
-	-	-	ضرب	ي	-	يضرِب	yuC ₁ AC ₂ iC ₃ u
هم	-	-	ضرب	ي	-	يضرِبهم	yaC ₁ C ₂ iC ₃ uhum
هم	-	-	ضرب	ي	س	سيضرِبهم	syaC ₁ C ₂ iC ₃ uhum
هم	-	-	ضرب	ي	س	سيضرِبها	syuC ₁ AC ₂ iC ₃ uhum
هم	ون	-	ضرب	ي	س	سيضرِبونها	syuC ₁ AC ₂ iC ₃ unahum

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]

Suffix	Suffix description	Suffix category Verb/Noun/Both	Example	phonetics
ني	First person	Verb	ضربني	drbny
ك	Second	Both	ضربك	drbk
هـ	Third person	Both	ضربه	drbh
ها	Third person	Both	ضربها	drbha
هم	Third person	Both	ضربهم	drbhm
هن	Third person	Both	ضربهن	drbhn
هما	Third person	Both	ضربهما	drbhma
كم	Second	Both	ضربكم	drbkm
كن	Second	Both	ضربكن	drbkn
كما	Second	Both	ضربكما	drbkma

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

Group #	Form #	Structure	Arabic imperfect	Arabic perfect
Group I	Form I	yaC ₁ C ₂ aC ₃	يَفْعَلُ	فَعَلَ
Group II	Form II	yuC ₁ aC ₂ C ₂ iC ₃	يُفْعِلُ	فَعَّلَ
	Form III	yuC ₁ aC ₂ iC ₃	يُفَاعِلُ	فَاعَلَ
	Form IV	yuC ₁ C ₂ iC ₃	يُفْعِلُ	فَعَّلَ
Group III	Form V	ytaC ₁ aC ₂ C ₂ aC ₃	يَتَفَعَّلُ	تَفَعَّلَ
	Form VI	ytaC ₁ AC ₂ aC ₃	يَتَفَاعَلُ	تَفَاعَلَ
Group IV	Form VII	yanC ₁ aC ₂ iC ₃	يَنْفَعِلُ	انْفَعَلَ
	Form VIII	yaC ₁ taC ₂ iC ₃	يَنْفَعِلُ	انْفَعَلَ
	Form X	ystaC ₁ C ₂ iC ₃	يَسْتَفَعِلُ	اسْتَفَعَلَ

- 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/ ال girls/ بنات saw/ رأت the/ ال paper/ ورقة/

(girls/ plural, feminine, and 3rd person).

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)*

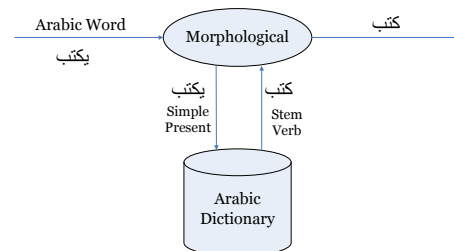


Figure 2. Morphological Analysis with process 5

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*

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 رأت (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 produced 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 رأت البنات الورقة "ra?at albnat alwaraqah" in the case of using VSO or رأينَ البنات الورقة "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.
9. Translation of a preposition.
10. Conjunction with "and
11. Adding the preposition " ال".
12. addition and deletion.

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 4. Experiment results

No.	ALKAF I	GOOGLE	TARJIM	Our System
Percentage	92.1%	84.6%	94.2	96.1%

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

- [1] Abu-Ata B., "An Arabic stemming algorithm on ERA for information retrieval", PhD. Thesis, Universiti Kebangsaan Malaysia, 2001.
- [2] Abu Shqeer O., "Handling Agreement and Word Reordering in English to Arabic Machine Translation", Master Thesis submitted to University of Sains Malaysia, 2002.
- [3] Abu Shquier M. and Sembok T., "Word Agreement and ordering In English-Arabic machine translation", Proceeding of the International Symposium on Information Technology, 2008.

- [4] Abu Shquier, M. & Abu Shqeer, O. (2012). Words ordering and corresponding verb-subject agreements in English-Arabic machine translation: Hybrid-based approach. *International Journal of Soft Computing And Software Engineering (JSCSE)*, 49-60.
- [5] Al-Dahdah, A. 1985. Arabic Language Grammar Dictionary. "معجم قواعد اللغة العربية". Beirut: Lebanon. Lebanon Library.
- [6] Al-Fedaghi, S.S., & Al-Sadoun, H.B. 1990. Morphological Compression of Arabic Text. *Information Processing & Management*. 26(2): 303-316.
- [7] Al-Najem, S., Inheritance-based Approach to Arabic Verbal Root-and-Pattern Morphology
- [8] Al-Jarf R., "SVO Word Order Errors in English-Arabic Translation", *Translators' Journal*, vol. 52, 2007, p. 299-308, 2007
- [9] Al-Khuli, M., (1982): Al-Taraakeeb al-ahaa?i9a fi llugha al-Arabiyya (Common Structures in Arabic), Dar Al-Uloom Publishing Co.
- [10] Al-Momani, I. 2010. Does The VP Node Exist In Modern Standard Arabic? *Journal of Language and Literature*, ISSN: 2078-0303, May, 2010.
- [11] Anderson, Stephen and Sandra Chung. 1977. On grammatical relations and clause structure in verb initial languages. In *Grammatical Relations: Syntax and Semantics 8*, eds. Peter Cole and Jerrold Sadock. 1-25. New York: Academic Press.
- [12] Attia M., 2008. "Handling Arabic Morphological and Syntactic Ambiguity within the LFG Framework with a View to Machine Translation", PhD Thesis, University of Manchester. Faculty of Humanities.
- [13] Badawi, E., Carter, M.G. & Gully, A. (2004). *Modern Written Arabic: A Comprehensive Grammar*. New York: Routledge.
- [14] Belkredim., F. a. (2009). An Ontology Based Formalism for the Arabic Language Using Verbs and their Derivatives. *Communications of the IBIMA Volume 11*. ISSN: 1943-7765.
- [15] Corbett G. "Agreement: Terms and boundaries. In William Griffin", *The Role of Agreement in Natural Language*, Proceedings of Texas Linguistic Society Conference, Austin, Texas, 2001.
- [16] El-Sadany T., and Hashish M., "An Arabic morphological system". *IBM Systems Journal* 28 (4): 600-612., 1989.
- [17] Feras Saeed. 2010 *Word Order and Agreement Asymmetry (English and Foreign Languages University, Hyderabad)*. The 5th annual Moscow Student conference on Linguistics
- [18] Hatem A. and Omer N., "Syntactic Reordering for Arabic- English Phrase-Based Machine Translation", 2010.
- [19] Hendrick, R., 2000. Celtic initials. In *The Syntax of Verb-Initial Languages*, eds. Andrew Carnie and Eithne Guilfoyle. 13-38. Oxford: Oxford University Press.
- [20] K., R. (2005). *Reference Grammar of Modern Standard Arabic*. Cambridge University Press, New York, USA.
- [21] Kramer, R. (2009), VSO and SVO word order in Middle Egyptian. In *Afroasiatic Studies in Memory of Robert Hetzron*, ed. C. Häberl. Cambridge: Cambridge Scholars Press. 92-147.
- [22] Mohammad, M.A., 2000. *Word Order, Agreement and Pronominalization in Standard and Palestinian Arabic*. Amsterdam: John Benjamins.
- [23] Ryding K., "A Reference Grammar of Modern Standard Arabic. Cambridge": Cambridge University Press, 2005
- [24] Saliba B., and Al-Dannan A., "Automatic morphological analysis of Arabic: a study of content word analysis", *Proc. of the First Kuwait Computer conference*. pp. 231-243., 1990.
- [25] Shaalan, K. (2005). *Arabic GramCheck: A grammar. Software Practice and Experience*, John Wiley & sons Ltd., 35(7):643-665.
- [26] Shaalan. K., a. A. (2009). *Syntactic Generation of Arabic in Interlingua-based Machine Translation Framework*. Third Workshop on Computational Approaches to Arabic Script-based Languages.

- [27] Sharif, M. (1993): Al-Tarakeeb al-naHwiyya (Grammatical Structures), Al-Shabaab Bookstore Publishing.
- [28] Sproat, Richard. 1985. Welsh syntax and VSO structure. *Natural Language and Linguistic Theory* 3: 173-216.
- [29] Wightwick, J. & Gaffar, M. (2008). *Arabic Verbs and Essentials of Grammar*. New York: McGraw-Hill.
- [30] Wickens G., "Arabic Grammar", Cambridge: Cambridge University Press, 1980.
- [31] Woolford, Ellen. 1991. VP-Internal Subjects in VSO and nonconfigurational languages. *Linguistic Inquiry* 22: 503-540.
- [32] Yusif J., "Automatic Part Of Speech Tagger for Arabic Language using Neural Network". PhD Thesis submitted to National University of Malaysia, 2007.