An Approach for Analyzing and Correcting Spelling Errors for Non-native Arabic learners

Khaled Shaalan#1, Rana Aref*2, Aly Fahmy*3

#Faculty of Informatics,British University in Dubai
Dubai International Academic City, PO Box 345015, Dubai, United Arab Emirates
1 khaled.shaalan@buid.ac.ae
*Department of Computer Science, Faculty of Computers and Information,Cairo University
5 Ahmed Zewel St.,PO Box 12613,Giza, Egypt
2 r.aref@fci-cu.edu.eg
3 a.fahmy@fci-cu.edu.eg

***Abstract*—Spellcheckers are widely used in many software products for identifying errors in users' writings. However, they are not designed to address spelling errors made by non-native learners of a language. As a matter of fact, spelling errors made by non-native learners are more than just misspellings. Non-native learners' errors require special handling in terms of detection and correction, especially when it comes to morphologically rich languages such as Arabic, which have few related resources. In this paper, we address common error patterns made by non-native Arabic learners and suggest a two-layer spell-checking approach, including spelling error detection and correction. The proposed error detection mechanism is applied on top of Buckwalter's Arabic morphological analyzer in order to demonstrate the capability of our approach in detecting possible spelling errors. The correction mechanism adopts a rule-based edit distance algorithm. Rules are designed in accordance with common spelling error patterns made by Arabic learners. Error correction uses a multiple filtering mechanism to propose final corrections. The approach utilizes semantic information given in exercising questions in order to achieve highly accurate detection and correction of spelling errors made by non-native Arabic learners. Finally, the proposed approach was evaluated using real test data and promising results were achieved.**

***Keywords*— Spelling Error Detection, Spelling Error Correction, Non-native Arabic Learners**

I. Introduction

Words are a core language element used for communicating within a language. Inefficient use of words of a language strongly affects communicating with this language. As a result, learners of second languages need to build a strong vocabulary for the language being learned and know how to effectively use this vocabulary in communication. Therefore, a spell checker plays an important role in developing new learners' vocabulary through detecting and correcting their common spelling mistakes [1]. Although spell checkers are widely available for a number of languages, including Arabic [2] [3] [4] [5] [6] [7] [8] [9], yet they are geared towards native speakers and have many shortcomings that prevent them from being efficient enough for non-native language learners. Mostly, native speakers' errors patterns are fixed; these errors can be typographical mistakes and most frequently involve one mistake per word [10]. On the other hand, non-native learners' errors involve multiple errors, occurring in any place within a word and proved to have significant deviation from the target word [10]. Most of the available checkers only detect errors and propose corrections regardless whether a true diagnosis or sufficient feedback would promote language learning. In addition, correcting errors regardless of their context highly increases ambiguity and may result in totally incorrect corrections for a given word [11]. In this paper, we propose a two-layer spell-checking approach for diagnosing and correcting spelling errors made by second language learners of Arabic. The proposed spelling detection mechanism is applied on top of the Buckwalter's Arabic morphological analyzer to identify possible spelling errors. The correction mechanism adopts a rule-based edit distance algorithm based on transformation rules designed in accordance with common errors made by Arabic learners. Error correction uses a multiple filtering mechanism to propose final corrections list. Moreover, spelling detection and correction is semantically-sensitive and based on semantic linguistic information given in exercising questions. The paper is structured as follows. Section 2 introduces an overview for the related work on spell checking and correction techniques. Section 3 describes the representation of questions used for testing the proposed approach. Section 4 introduces the error typology used for error detection and correction. The proposed detection and correction algorithms are described in Section 5. Evaluation results are presented in Section 6. Section 7 gives some concluding remarks.

II. Related Work

Spell checkers designed for Arabic language increased dramatically in the last decade due to the increased demand for Arabic applications requiring spell checking and correction facilities. Few Arabic spell checkers are now available for word processing applications, including: 1) commercial applications, like Microsoft word spell checker for Arabic [2], and 2) free office applications, like OpenOffice.org that includes Ayaspell spell checker [5]. Moreover, many Arabic portals are available in addition to a number of search engines available for Arabic which implements internal spell checkers [4] [6] [7] [8] [9]. Several research studies shed the light on the development of Intelligent Language Tutoring Systems (ILTS) for Arabic, for example, [12] [13]. However, theses systems as well as current Arabic spelling checker systems lack the focus on issues related to the learning aspects gained from lexical errors diagnosis and correction. As a result, these systems are considered insufficient for Arabic language learners. It is worth noting that spell checkers targeting non-native Arabic learners are rarely considered in the literature [12], although similar systems are available for other language [14] [1] [15].

A. Spelling Error Detection Techniques

Spelling error detection is concerned with detecting a word as an incorrect word. Techniques used for spell checking in many applications include dictionary lookup techniques, which are widely used, where words are compared and located in a language dictionary. Failing to detect a word in a dictionary indicates a spelling error. A commonly used technique is N-grams algorithms [16]. An n-gram is an N letter subsequence of a string; where N is usually 1, 2, or 3 [17]. N-gram analysis is used to estimate whether each N-gram encountered in an input string representing some text, is likely to be valid in the language [16]. Other techniques used for spell detection include, two-level morphological analysis as in Turkish of Oflazer (1996) [18], finite-state transducers [1], and machine learning algorithms [19]. Nevertheless, hybrid techniques are used also to perform spell checking [17]

B. Spelling Error Correction Techniques

Techniques for spelling errors correction can be categorized into context-dependent and context-independent error corrections [16]. Context-dependent error correction performs correction according to contextual or linguistic information available. On the contrary, context-independent correction performs correction to isolated words regardless of any contextual or linguistic information. By far, the most common studied and used technique for error correction in a number of languages is edit distance [16] [20]. Edit distance technique operates by computing the minimum number of editing operations (i.e., insertions, deletions, and substitutions) required to transform misspelled words into a valid dictionary entry [20]. Other techniques explored for context independent correction include: Similarity key, rule-based, N-gram, probabilistic, and neural network techniques. A similarity key technique [16] [20] is designed to transform words into keys, such that similarly spelled words will have similar keys. Hence computing keys for misspelled words will point to similarly spelled words in the lexicon. A rule-based technique [16] [20] involves algorithms to represent common spelling error patterns in the form of rules which are used to transform misspelled words into correct ones. An N-gram based technique (described in Section 2.1) is either used as a standalone or in conjunction with other techniques to perform error correction. A probabilistic technique [20] based on N-gram technique. A neural network [14] technique capable of doing associative recall based on incomplete input [16].

Context dependent error correction relies on contextual information available to detect and correct errors. Related error corrections are commonly used for detecting real word errors, which results from using valid language words in the wrong context. A number of techniques address this issue including: confusion sets, machine learning, phonetic similarity, semantic distance, and neural networks techniques. Approaches based on confusion sets, where confusion set includes common dictionary words that are commonly used in place of one another [11]. Machine learning techniques [21] are used in conjunction with confusion sets. Phonetic similarity techniques [22] are based on common phonological errors ensuing from incorrect spelling of pronunciation. Phonetic similarity techniques are commonly used for database query and search engines spelling corrections. SOUNDEX [9], PHONIX [22] and Metaphone Double [23] are all based on phonetic similarity techniques. Semantic distance techniques [22] are based on comparing word semantic with surrounding words. Moreover, techniques based on neural networks proved to offer good results for spelling corrections [22]

All pre-mentioned techniques address the issue of spelling corrections in a number of languages in concern with native language users. However, few techniques are proposed for spelling correction to non-native language learners in a number of languages [14] [1] [15]. Yet, no available systems address spelling correction for non-native Arabic learners.

III. Error Typology

In order to propose an efficient spelling correction algorithm for non-native Arabic learners, we significantly need to study and categorize common error patterns made by these learners. There are a number of resources [24] [25] [12] [26] that inspired our classification of common spelling errors made by Arabic learners. Common error patterns made by non-native Arabic learners are categorized as follows:

- *Editing Errors*: These errors result when learners misspell a word by omitting, adding, replacing or duplicating a letter within a given word. For example, spelling the word ' متبة' instead of the 'مكتبة'/library, 'مطببعية' instead of 'مطبعة'/printer,'شجره' instead of 'شجرة'/tree and 'كتتاب' instead of 'كتاب'/book
- *Phonetic Errors*: These errors result from recognizing and writing a letter instead of another having a similar pronunciation. Usually, these types of errors happen when users fail to map a specific phoneme of a certain word into a grapheme. This type of error is common and peculiar to Arabic as usually other languages do not have such closer similarities in pronunciation between letters. For example, spelling the word 'يحضر'/ attends as 'يحدر'. The letter 'ض' in the word 'يحضر'/attends is mistakenly exchanged with the letter 'د' which have almost similar pronunciation. Phonemes with close pronunciation are grouped in categories according to studies by [27] and [26], as shown in Table 1.
- *Vowel Errors*: These errors result from exchanging long vowels with short ones or vice versa. For example, the word نمتُ/slept-I ends with a damma diacritic. Spelling this word as نمتو is considered a vowel error.

- *Tanween Errors*: These errors result when learners misinterpret the sound of tanween pronounced at the end of a word, by adding an extra 'ن' letter. Tanween errors are considered a special case of vowel errors; yet they require special handling as they have a number of special cases. For example, spelling the word 'أكلا'/food in the sentence 'أكلت أكلا'/I ate food, as 'أكلن' is encountered as a tanween error
- *Shadda Errors*: These errors result when learners interpret Shadda diacritic as two similar letters instead of a single letter with shadda diacritic. For example, spelling 'دراجة'/bicycle as 'درراجة' is encountered as a shadda error.
- *Semantic Spelling Errors*: These errors result when learners use correct Arabic words in a wrong context within a sentence [25]. For example, 'ذهب إلي الكتب'/He went to the book. The word 'الكتاب'/book is a valid Arabic word but doesn't concur with the meaning of the sentence

In accordance with this common spelling errors classification, our error detection and correction approach is described in details in the next section

TABLE I: Phonological Classes

Phonological Class	Example
ث س ص	Spelling 'عصفور'/bird as 'عسفور'
أ يء وء	Spelling 'موءثرة'/impressive as 'مواثرة'
د ط ت ض	Spelling 'دجاجة'/chicken as 'ضجاجة'
ز ظ	Spelling 'زرت'/visited as 'ظرت'
خ غ	Spelling 'صغير'/young as 'صخير'
ت ة	Spelling 'دجاجة'/chicken as 'دجاجت'
ه ة	Spelling 'شجرة'/tree as 'شجره'
ق ك	Spelling 'قام'/undertake as 'كام'
ي يء	Spelling 'جريء' as 'جري'
و وء	Spelling 'موءثرة'/impressive as 'موثرة'

IV. The Proposed Approach

Our proposed spelling detection and correction approach follows a two-layer approach, including spelling error detection and correction. The approach is based on a questioning environment incorporating exercises designed for new learners of Arabic.

A. Questioning Environment

Questions based exercises are generated from an existing set of questions designed to address learners of Arabic. Learners are then asked to freely supply their answers to different questions. Supplied answers are checked to detect common spelling errors and propose corrections with descriptive feedbacks for causes of detected errors.

Questions are classified into two types: picture for spelling and spelling with sound questions. The first type displays a picture to the learner and asks him/her to write an Arabic word that describes the object in the picture, e.g., ship, apple, etc. This type of spelling question measures the learner's vocabulary knowledge and writing skills. The second type of spelling questions plays the sound of an Arabic word and the learner is asked to write down a word describing the segment s/he heard. This type of spelling questions measures the learner's listening and writing skills addressing possible phonological or pronunciation errors that most novice Arabic learners make.

All questions include semantic information in order to accurately perform contextual error correction. Each question is associated with an English glossary to refer to the object in question.

B. Spelling Error Detection

Detecting a spelling error involves detecting two types of errors: 1) ill-formed word errors and, 2) semantically incorrect errors.

Detecting ill-formed Arabic words is considered straightforward; it is mainly concerned with detecting syntactic errors in learners' answer. A semantically incorrect error is related to well-formed Arabic word input that deviates from the requirements of questions. It is detected by comparing the semantic features of the learner input with the semantic features stored with the question.

For detecting ill-formed Arabic words we utilize Buckwalter's Arabic Morphological Analyzer (BAMA), which is available as a free open source [28]. We used BAMA, Version 1, which includes 38,600 lemmas. It also provides multiple analyses of input Arabic words as well as an English glossary for each analyzed word. BAMA is used for analyzing well-formed Arabic words. Failing to analyze a word would indicate a possible ill-formed input error is very likely to be incompatible with Arabic morphological rules. On the contrary, successfully by passing BAMA analysis ensures no spelling mistakes in learner's input, yet this doesn't assure inexistence of semantic errors that are not compatible with question requirements. To detect semantically incorrect errors, all input word interpretations, generated by BAMA, are compared with given semantic features of the question being answered. Failing to match with a target meaning detects a semantically incorrect error.

C. Spelling Error Correction

Spelling error detection is ensued by spelling errors correction to generate possible word corrections for learners. Possibly detected errors, either ill-formed word or semantically incorrect errors, follow the same correction mechanism to propose a valid correction. Our proposed spelling correction approach for non-native Arabic language adopts the edit distance technique in conjunction with rule-based transformation approach.

Applying edit distance algorithm can generate possible corrections through editing errors made by learners. However, it is considered insufficient as a standalone mechanism for suggesting spelling correction since it gives no concern about common irregular error patterns made by non-native Arabic learners, as mentioned in Section 3. Consequently, we further apply a heuristic rule-based transformation approach to convert a misspelled Arabic word into a possible word correction.

Valid words obtained are used to generate an informative correction list more tailored to Arabic learners. In advance, multiple rules are applied to detect more complicated error structures. Every transformation rule applied to a word is logged to a logger attached to this word. Logging information can be used for generating appropriate diagnosing and feedback messages for learners. This will be described in details in a forthcoming publication. Table 2 describes transformation rules to be applied for each spelling error type.

For correcting the misspelled input word W, we apply the following steps:

1) Edit(Wji): Apply Edit distance algorithm to the erroneous word W; where i presents the letter being transformed and j represents the editing operation to be performed. Possible editing operations include letter omission, letter addition, letter substitution, and letter reversing.
2) Transform(W): Apply transformations rules to W.
3) ACC(CWL): Add generated transformed words to the correction word list CWL
4) Detect_Mutiple_Error(CWL): Reapply the algorithm to CWL to detect multiple errors, if any
5) Filter(CWL): Apply filtration mechanisms to CWL to generate the final CWL, Section 4.4 describes the filtering mechanism in details

Usually, the number of proposed corrections can overgenerate and increase exponentially. Thus, we propose a filtering mechanism to remove irrelevant generations from final correction word list to yield possible word correction for learners. Our proposed filtering methods are designed to minimize, to a large extent, the generated correction word list. We designed our filtering algorithm to be adaptive, beginning with a wide acceptance interval and tighten up the filter as better candidates appear. Moreover, the filter excludes corrections that do not concur in context with the question being answered. Filtering mechanisms are explained in details in the next section.

D. Filtering

We designed two types of filters to extensively minimize the generated correction word list for learner's erroneous input, namely: Morphological analysis filtering and gloss filtering.

1) Morphological Analyzer Filter: Applying the correction process does not confirm that all words in the correction word list are valid Arabic words. Hence, we apply morphological analysis filtering to filter out all non Arabic words. Morphological analysis filtering implies sending each word in the correction word list to BAMA to generate an interpretation list including an English glossary for each word. This glossary is to be used in gloss filtering in correspondence with questions' requirements. For example, applying a vowel transformation function to the word 'حائيط' which is a misspelled form of the word 'حائط'/wall results in the word 'حئط' which is not a genuine Arabic word and hence should be excluded from the correction word list.

2) Gloss Filter: Gloss filters are further applied to adapted correction word list generated by the morphological analysis filter. Gloss filtering omits all possible word corrections with a context different from that expected for the question being answered. Accordingly, gloss filtering can extensively minimize the correction list by excluding irrelevant and ambiguous transformations/input. For example, if a spelling question displays a happy face to a learner and asks him/her to write an adjective, i.e. a single word, which describes this picture, see Fig. 1. The expected solution should have the meaning of being happy ('سعيد').
A possible learner's solution with a short to long vowel error is 'ساعيد' (the second letter 'ا' is a long vowel that is incorrectly replaced by the short vowel Fatha) applying vowel transformations to this word results in the following possible word corrections: 'سعيد'/happy and 'ساعد'/helped. Both words are considered valid Arabic words and have the interpretations shown in Appendix I, two interpretations for 'سعيد'/happy and nine interpretations for 'ساعد'/helped, that were generated by BAMA during morphological filtering. Applying a gloss filter to proposed corrections for word 'ساعيد', excludes all the nine solutions for word 'ساعد' since none of them is glossed as "being happy". This minimizes the correction word list by eliminating corrections with different semantics. Moreover, the gloss filter minimizes generated solutions for the word 'سعيد' by excluding the second solution since it is glossed as said/saeed which does not mean happy. This leaves us with only the first solution of the correction 'سعيد' and since it is glossed as "happy", which matches with the expected glossary for the correct answer, it will be reported to the learner as corrected form of his semantically incorrect input.

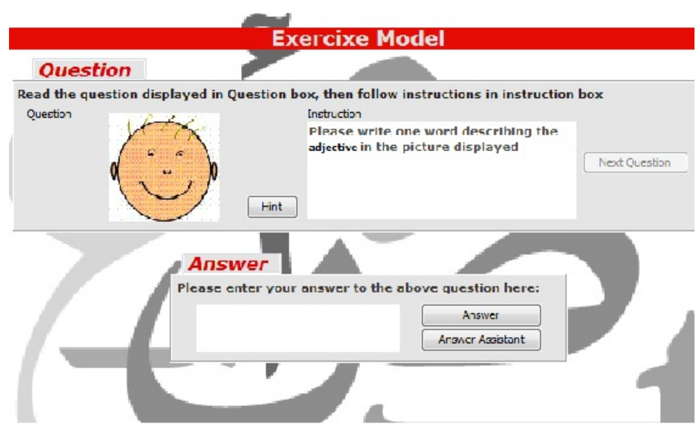

Fig. 1: Spelling Question

V. Evaluation

Our evaluation methodology is based on the common natural language processing (NLP) measures: precision and recall. We evaluated our proposed spelling detection and correction approach using a set of test data composed of 190 misspelled words. Test data include both single and multi-error misspellings composed of up to three errors per word. Average word length is 5 letters per word. Test data are collected from an error corpora [5] acquired from related studies [28] [25] [12]and actual learner's errors [12]. Test data is designed to cover all types of spelling errors mentioned in this research. The test data is related to a

TABLE II: Spelling Errors Transformations Rules

ErrorType	*Transformationrule*
Phonemes Exchange	Letters occurring within the same phonological class, see Table 1, are exchanged to generate possible corrections list. For example, the letter 'ز' in the misspelled word 'زيل' should be exchanged with 'ذ' to get the correct word 'ذيل'/tail
Vowel Errors	Long vowels in learners' inputs are replaced by short ones and the vice versa to propose a correction. For example, in the misspelled word 'نمتو' the letter 'و' is exchanged with damma diacritic, the resulted word 'نمتُ'/I slept is considered a valid correction
Tanween Errors	Misspelled words ending in a 'ن' letter are transformed by removing: • 'ون' suffix letters if any • 'ين' suffix letters if any • 'ان' suffix letters if any • 'ن' letter if not preceded by 'ي', 'و', or 'ا'. Removing letters to correct a misspelled word may involve addition of letters. For example, in case of fatha tanween the letter 'ا' should be added to the end of the word. Consider the word 'نورن', it is transformed by removing the letter 'ن' and adding the letter 'ا' yielding the correct word 'نورَاً'/light This transformation is not straightforward and needs special handling for another related cases: • If a word ends with any hamza letter, for example 'أ ي ء و ء', the extra 'ا' letter is not added at the end of the word in case of Fatha Tanween and the occurrence of an extra 'ا' letter before 'ن' letter is removed. For example a tanween transformation for the word 'رداءن' is 'رِدَاءً'/robe • If a word ends with 'ة' Taa Marbouta letter the sound of 'ت' Taa Maftouha letter results when adding tanween. Applying tanween transformation to words involves removing the invalid letter 'ت' if exists. For example applying this transformation to the word 'شجرتين' results in the correct word 'شجرةٍ'/tree • If extra 'ا' letter precedes the Tanween letter it should be eliminated. For example, applying tanween transformation to the word 'كتاباان' results in the correct 'كتابا'/book • If a word ends with 'ى' Alf Maksoura no addition of the 'ا' letter is performed for Fatha Tanween but instead a 'ى' letter could be added. For example, applying tanween transformation to the word 'فتان' results in the correct word 'فتى'/boy
Shadda Errors	The doubled letter is removed generating possible word corrections. For example applying the shadda transformation to the word 'الضالLين' with doubled 'ل' letter results in the correct word 'الضالين'/misled

set of 315 question, 175 sound segment questions, and 140 picture questions. All questions include relevant semantic information. In correspondence with the proposed spelling error detection and correction, if a spelling error is successfully detected or corrected, this indicates a true positive detection and correction; whereas, undetectable errors or uncorrected errors indicate false negative detection or correction. While detecting or correcting a non erroneous input indicates a false positive detection or correction. Table 3 gives examples for different cases of spelling errors detection and correction. In order to measure the performance of our proposed approach, we calculated precision and recall rates for both spelling error detection and correction. Precision (P) and Recall (R) rates are calculated [21] as follows

$$R = \frac{True\_positives}{True\_positives + False\_negatives} \quad (1)$$

$$P = \frac{True\_positives}{True\_positives + False\_positives} \quad (2)$$

Applying the test data to questions, based on exercises, using our error detection and correction developed system, the results shown in Table 4 were achieved for each type of spelling error. An 80+% recall were achieved and a 90+% precision were achieved.

VI. Conclusion

In this paper, we presented and evaluated a proposed approach for spelling error detection and correction targeted at non-native Arabic language learners. We utilized Buckwalter's morphological analyzer for detection of spelling errors. We further applied edit distance algorithm in addition to rule-based transformation approach. Rules are formulated according to our study on common spelling errors made by Arabic learners. Applying filtering mechanisms for spelling corrections, specially gloss related filters proved to dramatically decrease the proposed correction lists. Results of the proposed evaluation methodology in an exercising based environment clearly proved the success of the proposed two-layer spelling error detection and correction approach, especially that are related to detecting and correcting semantic based spelling errors. According to spelling correction precision and recall rates, adding semantic information proved to significantly affect the proposed approach. Evaluating our technique in an exercise based environment certify that it can be successfully integrated with intelligent language tutoring systems for Arabic to include advanced lexical errors considerations.

References

[1] S. L'haire, "Fipsortho: A spell checker for learners of French", *ReCALL*, vol. 19, n. 3, pp. 137–161, May 2007.
[2] "Microsoft Products and Arabic Support", http://www.microsoft.com/middleeast/arabicdev.

TABLE III: Spelling detection and correction examples

Example	Input word	Proposed Correction	Intended word
True positive detection and correction	ويساضة	وسادة/pillow	وسادة/pillow
False positive detection	أصبع	–	باهم/thumb
True positive detection and false positive correction	شامال	شمالي/ northern	شمال/north
False negative correction	أوت	Not corrected	أسود/black

TABLE IV: Evaluation results

Error Type	Detection		Correction	
	Recall	Precision	Recall	Precision
Phonological Errors	0.9433962	0.909090	0.8333333	0.9433962
Vowel Errors	0.9850746	0.9677419	0.825	0.9850746
Tanween errors	0.8620689	0.9803921	0.875	0.9655172
Shadda Errors	0.9756097	0.9999999	0.9090909	0.9999999
Editing Errors	0.95238095	0.95238095	0.8695652	0.9090909
Semantic errors	0.9999999	0.9999999	0.95	0.9047619

[3] "Aspell", http://arabic-spell.sourceforge.net/.
[4] Sakhr, "Al Idrissi", www.sakhr.com.
[5] "Ayaspell project", http://ayaspell.sourceforge.net/.
[6] "Ayna", http://www.ayna.com/.
[7] Google, "Google Search Engine", http://www.google.com.
[8] Konouz, "Konouz Search Engine", http://www.konouz.com.
[9] Wikipedia, "Soundex", http://en.wikipedia.org/wiki/Soundex.
[10] Hovermale D.J. and Dennis N. Mehay, "Real-word Spelling Correction for CALL,", 2009, Presentation for The Sixth Midwest Computational Linguistics Colloquium.
[11] A. Carlson and I. Fette, "Memory-Based Context-Sensitive Spelling Correction at Web Scale", *ICMLA*, pp. 166–171, 2007.
[12] Magdy Marwa, "Lexical Error Diagnosis for Second Language Learners of Arabic", Master's thesis, Department of Computer Science, Faculty of Computer and Information, Cairo University, Cairo,Egypt, 2008.
[13] K. Shaalan, "An Intelligent Computer Assisted Language Learning System for Arabic Learners In Computer Assisted Language Learning", *An International Journal, Taylor Francis Group Ltd.*, vol. 18, n. 1-2, pp. 81–108, 2005.
[14] Hovermale D. J., "SCALE: Spelling correction adapted for learners of English", in *Proceedings of the CALICO-08 ICALL Special Interest Group Pre-conference Workshop*, San Francisco, CA, USA, 2008.
[15] A. Rimrott, "SANTY: A Spell Checking Algorithm for Treating Predictable Verb Inflection Mistakes Made by Non-Native Writers of German", *LING 807 Computational Linguistics, Simon Fraser University*, 2003, Term Paper.
[16] Bruce W. Watson Hsuan Lorraine Liang and Derrick G. Kourie, *Classification for Selected Spell Checkers and Correctors*, School of Computing, University of South Africa.
[17] JIANG Xue-mei XIE Fang, "Error analysis and the EFL classroom teaching", *US-China Education Review*, vol. 4, n. 9, pp. 10–14, 2007.
[18] A. Solak, "Design and implementation of a spelling checker for Turkish", Master's thesis, Department of Computer Engineering and Information Sciences, Bilkent University, Ankara, Turkiye, April 1991.
[19] Ola Knutsson Jonas Sjbergh, "Faking errors to avoid making errors: Machine learning for error detection in writing", in *Proceedings of RANLP 2005*, 2005.
[20] Karen Kukich, "Techniques for Automatically Correcting Words In Text", *ACM Computing Surveys*, vol. 24, n. 4, pp. 377–439, 1992.
[21] Aminul Islam and Diana Inkpen, "Real-Word Spelling Correction using Google Web 1T 3-grams", in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pp. 1241–1249, Singapore, 2009.
[22] S. Deorowicz and M.G. Ciura, "Correcting spelling errors by modeling their causes", *International Journal of Applied Mathematics and Computer Science*, vol. 15, n. 2, pp. 275–285, 2005.
[23] Wikipedia, "Double Metaphone", http://en.wikipedia.org/wiki/Double_Metaphone.
[24] Al-Anati W., "Noon Wal Qalam: Arabic for Speakers of Other Languages, An Educational Linguistic Study", *Umm AL-Qura University*, vol. 2, pp. 112–152, Jul 2009.
[25] J. A. Jassem, *Study on Second Language Learners of Arabic: An Error Analysis Approach*, Kuala Lumpur (Malaysia): A.S. Noordeen, 2000.
[26] Gomah A Shaalan K., Allam A., "Towards Automatic Spell Checking for Arabic", in *Proceedings of the 4th Conference on Language Engineering, Egyptian Society of Language Engineering (ELSE)*, pp. 240–247, Egypt, Oct 2003.
[27] S. Abdallah Antar, النظام الصوتي العربي و تعليم نطق الأصوات, *Project funded by CETL*, School of Oriental and African Studies, University of London.
[28] Aramorph, "Buckwalter's analyzer", http://www.nongnu.org/aramorph.

APPENDIX

Appendix I Morphological Interpretations of the word 'سعيد'/happy

```
SOLUTION #1
Lemma:saEiyd Vocalized as:saEiyd
Morphology :
prefix : Pref-0
stem : N/ap suffix : Suff-0
Grammatical category :
stem : saEiyd ADJ
Glossed as :
stem : happy

SOLUTION #2
Lemma:saEiyd Vocalized as:saEiyd
Morphology :
prefix : Pref-0
stem : N0 suffix : Suff-0
Grammatical category :
stem : saEiyd NOUN_PROP
Glossed as :
stem : Said/Saeed
```

Morphological Interpretations of the word 'ساعد'/helped

```
SOLUTION #1
Lemma:Ead˜ Vocalized as:sa>aEud˜
Morphology :
prefix : IVPref-AnA->a
stem : IV_V suffix : Suff-0
Grammatical category :
prefix : sa FUT prefix :>a IV1S
stem : Eud˜ VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem : count/consider/deem

SOLUTION #2
```

```
Lemma:EadA Vocalized as:sa>aEod
Morphology :
prefix : IVPref-AnA->a
stem : IV_0hwnyn suffix : Suff-0
Grammatical category :
prefix :sa FUT prefix : >a IV1S
stem : Eod VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem : attack/assail

SOLUTION #3
Lemma:waEaz Vocalized as:sa>aEid
Morphology :
prefix : IVPref-AnA->a
stem : IV suffix : Suff-0
Grammatical category :
prefix : sa FUT prefix: >a IV1S
stem : Eid VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem : promise

SOLUTION #4
Lemma:Ead~ Vocalized as:sa>uEad~
Morphology :
prefix : IVPref-AnA->u
stem : IV_V_Pass_yu suffix : Suff-0
Grammatical category :
prefix : sa FUT prefix : >u IV1S
stem : Ead~ VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem:be counted/be considered/be deemed

SOLUTION #5
Lemma:>aEad~ Vocalized as:sa>uEid~
Morphology :
prefix : IVPref-AnA->u
stem : IV_V_yu suffix : Suff-0
Grammatical category :
prefix : sa FUT prefix : >u IV1S
stem : Eid~ VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem : prepare/make ready

SOLUTION #6
Lemma:>aEad~ Vocalized as:sa>uEad~
Morphology :
prefix : IVPref-AnA->u
stem : IV_V_Pass_yu suffix :Suff-0
Grammatical category :
prefix : sa FUT prefix : >u IV1S
stem : Ead~ VERB_IMPERFECT
Glossed as :
prefix : will prefix : I
stem : be prepared/be made ready

SOLUTION #7
Lemma:sAEad Vocalized as:sAEada
Morphology :
prefix : Pref-0
stem : PV suffix:PVSuff-a
Grammatical category :
stem : sAEad VERB_PERFECT
suffix : a PVSUFF_SUBJ:3MS
Glossed as :
stem : help/assist/support/contribute
suffix : he/it <verb>

SOLUTION #8
Lemma:sAEad Vocalized as:sAEido
Morphology :
prefix : Pref-0
stem : CV suffix : CVSuff-o
Grammatical category :
stem : sAEid VERB_IMPERATIVE
suffix : o CVSUFF_SUBJ:2MS
Glossed as :
stem : help/assist suffix : you <verb>

SOLUTION #9
Lemma:sAEid Vocalized as:sAEid
Morphology :
prefix : Pref-0
stem : Ndu suffix : Suff-0
Grammatical category :
stem : sAEid NOUN
Glossed as :
stem : forearm
```