Arabic Error Feedback
in an Online Arabic Learning System

Khaled F. Shaalan[1,2] and Habib E. Talhami[1,2]

[1] Institute of Informatics
The British University in Dubai
P. O. Box 502216, Dubai, UAE
[2] Honorary Fellow, School of Informatics, University of Edinburgh
{khaled.shaalan,habib.talham}@buid.ac.ae

Abstract. Arabic is a Semitic language that is rich in its morphology and syntax. The very numerous and complex grammar rules of the language could be confusing even for Arabic native speakers. Many Arabic intelligent computer-assisted language-learning (ICALL) systems have neither deep error analysis nor sophisticated error handling. In this paper, we report an attempt at developing an error analyzer and error handler for Arabic as an important part of the Arabic ICALL system. In this system, the learners are encouraged to construct sentences freely in various contexts and are guided to recognize by themselves the errors or inappropriate usage of their language constructs. We used natural language processing (NLP) tools such as a morphological analyzer and a syntax analyzer for error analysis and to give feedback to the learner. Furthermore, we propose a mechanism of correction by the learner, which allows the learner to correct the typed sentence independently. This will result in the learner being able to figure out what the error is. Examples of error analysis and error handling will be given and will illustrate how the system works.

1 Introduction

Computer-assisted language learning (CALL) addresses the use of computers for language teaching and learning. CALL emerged in the early days of computers. Since the early 1960's, CALL systems have been designed and built. The effectiveness of CALL systems has been demonstrated by many researchers [6, 7]. More than a decade ago, Intelligent Computer-Assisted Language Learning (ICALL) started as a separate research field, when artificial intelligence (AI) technologies were mature enough to be included in language learning systems. The beginning of the new research field was characterized by intelligent tutoring systems (ITS), which embedded some NLP features to extend the functionality of traditional language learning systems. The continuous advances in ICALL systems have been documented in several publications [2, 3, 5, 9].

On of the weaknesses of current Arabic ICALL systems is that learners cannot key in Arabic sentences freely. Similarly, the system cannot guide the learner to correct the most likely ill-formed input sentences. The learner just accepts the information,

which has been pre-programmed into the system. For these systems to be useful, more research to combine NLP techniques with language learning systems is needed [8]. Parsing, the core component in ICALL systems, allows the system both to analyze the learner's input and to generate responses to that input [4]. Allowing learners to phrase their own sentences freely without following any pre-fixed rules can improve the effectiveness of ICALL systems, especially when the expected answers are relatively short and well-focused [1]. Both the well- and ill-formed structure of the input sentence can be recognized. The learner should be allowed to correct the typed sentence independently.

This paper describes error analysis and handling in an Arabic ICALL system using NLP techniques, which is a step towards enhancing current Arabic ICALL systems. The current system guides learners to recognize by themselves the errors or improper usage of their language constructs. In other words, it helps learners to learn from their own mistakes. It doesn't give them the correct answer directly but it enables them to try over and over again. In this system, we use NLP tools such as a morphological analyzer, a syntax analyzer, and an error analyzer to give feedback to the learner. Furthermore, we propose a mechanism of correction by learners which allows the learner to correct the typed sentence independently.

The rest of the paper is organized as follows: Arabic ICALL framework is summarized in Section 2. This is followed by a description of the Arabic sentence analysis in Section 3. Next, the proposed feedback component responsible for error analysis and error handling is described in Section 4. Finally, a conclusion and recommendations for further enhancements are given in Section 5.

2   The Arabic ICALL Framework

Figure 1 shows the overall framework of the proposed Arabic ICALL system by [8]. This system consists of the following components: user interface, course material, sentence analysis, and feedback. The user interface provides the means of communications between the learner and the Arabic ICALL system. The course material includes educational units, an item (question) bank, a test generator, and an acquisition tool. The sentence analysis includes a morphological analyzer, a syntax analyzer (parser), grammar rules, and a lexicon. The feedback component includes an error analyzer and error handler that are used to parse ill-formed learner input and to issue feedback to the learner.

3   Arabic Sentence Analysis

The sentence analysis in Arabic ICALL includes a morphological analyzer, a syntax analyzer (parser), grammar rules, and a lexicon. The sentence analysis works as follows. The learner written input is first fed into the interactive preprocessor, where it is grouped into words. The words in the input are then decomposed into roots and affixes by the morphological analyzer, which obtains information about the subparts from the lexicon (e.g., part of speech, number, case). The subparts so identified are
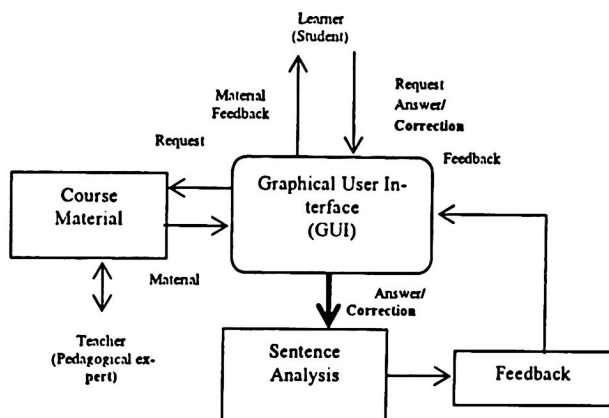
Learner
(Student)

Material
Feedback

Request
Answer/
Correction

Request

Feedback

Course
Material

Graphical User In-
terface
(GUI)

Material

Answer/
Correction

Teacher
(Pedagogical ex-
pert)

Sentence
Analysis

Feedback

Figure 1. Overall Framework of the Proposed Arabic ICALL System

then reunified into whole words and passed along to the syntactic parser. The Arabic parser, which is based on Definite Clause Grammar (DCG) formalism, tries to build a structure (usually, 'parse tree') based on the information from the lexicon concerning the grammatical relations between the words. The parser then applies a set of rules representing the grammar of Arabic until it finds the structure represented by the input sentence. This structure is passed to the feedback component that is equipped with an error analyzer and an error handler. The error analyzer identifies and records any errors made in the structural description which is generated. The error handler applies the error handling mechanism on the ill-formed leaner input to produce the appropriate feedback message.

3.1 The Grammar Formalism

Arabic grammar in Arabic ICALL is written in the DCG formalism. There two types of grammars that have been used for learning Arabic: Grammar rules for grammatically correct sentence and grammar rules for linguistic analysis (Iarab).

In the following, we show an excerpt of DCG rules used for parsing a grammatically correct Arabic verbal sentence.

```
(1)  verbal_sentence --> simple_verbal_sentence.
(2)  verbal_sentence --> prefixed_verbal_sentence.
(3)  verbal_sentence --> special_verbal_sentence.
(4)  simple_verbal_sentence --> verb, subject, object,
     unrestricted_object.
```

For simplicity, these rules do not include linguistic features such as gender, number and definition, which are assigned to each non-terminal. Rule (4) illustrates a grammar rule for parsing a simple verbal sentence that consists of four constituents: a verb, a subject, an object and an unrestricted object "قلطم لوغـم". An unrestricted object is a noun that originates from the infinitive verb. This kind of repetition is considered a

mark of good style. In Arabic, repeating the verbal noun after the verb makes the sentence more emphatic. This is explained by the following example:

<div dir="rtl">ساعدني مساعدة عظيمة</div>

He helped me a great deal of help

In the following, we show an excerpt of DCG rules used for the linguistic analysis of the words between brackets. These words are an object that is followed by an adjective such that the adjective agrees with the object (the noun it modifies) in number, gender, definition, and end case.

```
object(Words_bet_brackets, Rest, Analysis) -->
  [Object], [Adjective],
  {get_analysis(Object, Gender, Num, Def,
                Words_bet_brackets, Rest1, End_case,
                Analysis1, 'مفعول به'),
   get_analysis(Adjective, Gender, Num, Def, Rest1,
                Rest, End_case, Analysis2, 'نعت'),
   append(Analysis1, Analysis2, Analysis)
  }.
```

As an example, this rule can be used with the following question:

<div dir="rtl">أعرب ما بين القوسين في الجملة الآتية:</div>
<div dir="rtl">ألّف العقاد (كتبا كثيرة).</div>

Give the linguistic analysis of the words between brackets in the following sentence: Al-aakad authored (many books).

The parser produces a parse representing the linguistic analysis of the Arabic word that consists of a quadruple abstract representation of the canonical form:

<div dir="rtl">الموقع الإعرابي</div>	+	<div dir="rtl">الإعراب</div>	+	<div dir="rtl">علامة الإعراب</div>	+	<div dir="rtl">السبب</div>
Analytic location	+	End case	+	Analytic sign	+	Reason

4   The Feedback Component

Feedback is the computer's response to answers made by learners. We have augmented the Arabic grammar with heuristic rules (buggy rules) which are capable of parsing ill-formed input and which apply if the grammatical rules fail. The feedback system compares the analysis of the learner's answer with the correct answer that is generated by the system. If there is a match, a positive message will be sent to the learner. Otherwise, a feedback message will be sent to the learner. In the following subsections, we show how the system catches the learner's errors and how it handles the ill-formed natural language input.

4.1   Rules for Error Analysis

In our implementation, we have augmented the Arabic grammar with heuristic rules which are capable of parsing ill-formed input (buggy rules) and which apply if the

grammatical rules fail. As an example, consider the following question to complete a sentence with a suitable unrestricted object "المفعول المطلق":

<div dir="rtl">

بـسـانم قلطم لوعفـمب لمكأ:

_____ ىبأ ربأ .

</div>

Complete the following with the correct unrestricted object:
I am kind to my father _____.

The following is an analysis of the possible learner's answer along with the corresponding feedback:

- A word that is not a noun. Issue a message describing that the unrestricted object should be a noun.
- A word that is a noun but does not originate from the infinitive verb. Issue a message describing that the unrestricted object should be the infinitive of the verb.
- A word that is both a noun and originates from the infinitive verb but is defined. Issue a message describing that the unrestricted object should be undefined.
- A word that is a noun, originates from the infinitive verb but needs the end case "Alef Tanween", and is undefined. Issue a message describing that a missing end case of the unrestricted object.
- A Correct answer. Issue a positive message.

4.2   Error Handling Mechanism

Learner's responses which have special handling mechanisms in case of ill-formed learner input are: linguistic analysis, classification into categories, sentence transformation, and completing a sentence.

Handling of linguistic analysis. Linguistic analysis questions can apply either for an entire sentence or a part of it. The latter is usually a sequence of words between brackets. The following description outlines the steps for handling of linguistic analysis:

- Parse the given sentence (or the sequence of words between brackets) and generate its linguistic analysis in a quadruple abstract representation form
- Convert learner answer into the abstract representation form
- Compare the learner's answer with the generated answer to issue the appropriate feedback message

Example:

<div dir="rtl">

هل ترى فرق اقرا إعرابيا فيما بين القوسين؟
- استمتعت بجو الريف (استمتاعا).
- هذا إلى الريف (استمتاعا) بوجه.

</div>

What's the difference in linguistic analysis of the words in parentheses?
- *I enjoyed the country weather (very much)*
- *I go to the countryside (to enjoy) its weather*

The parser is used to analyze each of the input sentences. The generated correct linguistic analyses of the words in parentheses are the following:

- First word: ['مفعول مطلق', 'منصوب', 'فتحة', 'مفرد'] [unrestricted object, accusative, fat-hah, singular]
- Second word:['مفعول لأجله', 'منصوب', 'فتحة', 'مفرد'] [causative object, accusative, fat-hah, singular]

The difference, in this case, is in the analytic location (i.e. the first argument in the quadruple abstract representation). The learner's answer is also converted to the quadruple abstract representation. The comparison between these representations will issue the appropriate feedback that describes the source of the error. The possible source of the errors could be: incorrect analytic location, incorrect end case, incorrect reason, or a partially correct answer.

Handling of classification into categories. Classification into categories questions can apply either for identifying morphological categories or for identifying syntactic constituents (possibly, a complete sentence). The following description outlines the steps for handling classification into morphological categories:
- Morphologically analyze the words in the given sentence and determine the words features.
- Generate N lists, a classification of the words according to the questions words.
- Assign the learner answer to N Lists.
- Compare the learner's answer with the generated answer to issue the appropriate feedback message

Example:

عين الاسم والفعل والحرف في الجملة الآتية:
وقف التلاميذ احتراما للمعلم

Identify the category of each of the words in following sentence
The students stood up respecting the teacher

The morphological analyzer is used to analyze each inflected Arabic word to recognize its category. Then, according to the word category, the words are classified into three lists. The following is the generated correct answer:

- Verb: [['وقف',verb,male,singular,past,...]] (stood)
- Noun:[['المعلم',noun,male,singular,def,...], ['احتراما',noun, male,singular,undef, ...],
- ['التلاميذ',noun, male,plural,def, ...]] (the teacher, honoring, the students)
- Particle: [['ال',particle,def_article,...]] (the)

The learner's answer is also assigned to three lists containing verbs, nouns, and particles, respectively. The comparison between the corresponding lists will issue the appropriate feedback that describes the source of the error. The possible source of the errors could be: missing words from the respective morphological category, or assigning a word to an incorrect morphological category.

The following description outlines the steps for handling of classification into syntactic constituents:
- Parse the given sentence and determine the parse tree.
- Generate N lists, a classification of the sentence's constituents according to the questions words
- Assign the learner answer to N Lists.

Compare the learner's answer with the generated answer to issue the appropriate feedback message.

Example:

<div dir="rtl">
عيّن المبتدأ والخبر مبيّنا نوع العون في الجملة الآتية:
الجنود الشجعان ينتصرون
</div>

Identify the inchoative and enunciative, and the type of the enunciative in the following sentence.
the brave soldiers fought to victory

The parser is used to analyze the input sentence into a parse tree as follows:

```
nominal_sentence(
  inchoative(noun('الجنود',noun, male,plural,def,…),
             adj('الشجعان', noun,
  male,plural,def,adj,…)),
  enunciative(verbal_sentence(verb('ينتصرون',verb, male,
  plural,present,…)))
)
```

Then, according to the parse tree, the words are classified into three lists. The following is the generated correct answer:

-   inchoative: [noun('الجنود', noun, male,plural,def,…),adj('الشجعان', noun, male,plural,def,adj,…)] (the brave soldiers)
-   enunciative: [verbal_sentence(verb('ينتصرون', verb,male, plural,present,…))] (make victory)
-   enunciative type: [verbal_sentence]

The learner's answer is also assigned to three lists containing inchoative, enunciative, and enunciative type, respectively. The comparison between the corresponding lists will issue the appropriate feedback that describes the source of the error. The possible source of the errors could be: incorrect constituent type (analytic location), or assigning a syntactic constituent to an incorrect category.

Handling of transformation of a sentence. Transformation questions require the learner to change/rewrite the form of a sentence. The following description outlines the steps for handling of transformation of a sentence:

-   Parse the given sentence and determine the parse tree; apply a tree-to-tree transformation to generate the transformed parse tree
-   Parse the learner's answer to determine the parse tree.
-   Compare the parse tree of the learner's answer with the parse tree of the generated answer to issue the appropriate feedback message.

Example:

<div dir="rtl">
الجملة الآتية أسمية أجعلها فعلية:
المعلم يشرح الدرس
</div>

Change the following nominal sentence into verbal sentence:
The teacher teaches the lesson

The parser is used to analyze the input sentence into a parse tree as follows:

```
nominal_sentence(
  inchoative(noun('المعلم',noun,male,singular,def,…)),
  enunciative(verbal_sentence(verb('يشرح',verb,male,
  singular,present,…),object(noun('الدرس',noun,
  male,singular,def,…)))))
  )
```

Then, the parse tree of the nominal sentence is transformed to the following verbal sentence.

```
verbal_sentence(
  verb('يشرح',verb,male,singular,present,…),
  subject(noun('المعلم',noun,male,singular,def,…)),
  object(noun('الدرس', noun,male,singular,def,…))
  )
```

In addition, words of the transformed parse tree is grouped in a list as follows:

List of words: [verb('حرشي',verb,male,singular,present,…), noun('المعلم',
noun,male,singular,def,…),noun('الدرس', noun,male,singular,def,…)]

The learner's answer is also analyzed into a parse tree and words are grouped into a list. The comparison between the corresponding representations will issue the appropriate feedback that describes the source of the error. The possible source of errors could be: extra words, missing words, grammatically incorrect sentence, or incorrect transformation of a word (incorrect verb: tense, number, gender,…; incorrect noun: number, gender, definition,…).

Handling of Fill-in-the-blanks. Fill-in-the-blank questions can apply for the generation of isolated words with a particular form, or to the generation of words to complete a sentence that achieves feature agreement among its constituents. The following description outlines the steps for handling of rewriting of isolated words with particular morphological form:

- Morphologically generate the given words and determine their features.
- Morphologically analyze the learner's answer and determine the words features.
- Compare the parse tree of the learner's answer with the parse tree of the generated answer to issue the appropriate feedback message.

Example:

ضع جمع المثنى والجمع الكلمات التالية:

- المهندس
- الثمرة
- صحراء

What is the correct dual and regular plural of the following words?
- *Engineer*
- *Fruit*
- *Desert*

The morphological generator is used to synthesize the given words into the dual and plural forms taking into consideration the possible end case. The following is the generated correct answer:

- Dual list: [['المهندسان',noun,male,dual,def,nominal,...], [['المهندسين',noun,male, dual,def,accusative, ...], ['المثرتان', noun,female,dual,def,nominal, ...], ['المثرتين', noun,female,dual,def,accusative, ...], ['حصاوران', noun, female,dual, undef,nominal, ...],['حصاورين', noun, female,dual,undef, accusative, ...]]
- Plural list: [['المهندسون',noun,male,plural,def,nominal, ...], [['المهندسين', noun,male,plural,def,accusative, ...], ['المثرتات',noun,female,plural,def, ...], ['حصاورات', noun, female, plural,undef,...]]

The morphological analyzer is used to analyze each inflected Arabic word of the learner's answer. These words are classified into two lists containing dual and plural forms, respectively. The comparison between the corresponding lists will issue the appropriate feedback that describes the source of the error. The possible source of errors could be: different word, incorrect word category (switch dual forms with plural forms), incorrect generation of a word (incorrect verb: tense, number, gender,...; incorrect noun: number, gender, definition,...).

The following description outlines the steps for handling of fill-in-the-blank to achieve agreement among sentence constituents:

- Analyze the sentence and determine its constituents.
- Morphologically generate the missed words.
- Parse the learner's answer to determine the parse tree.
- Compare the parse tree of the learner's answer with the parse tree of the generated answer to issue the appropriate feedback message.

Example:

<div dir="rtl">

أكمل بمعنول مطلق مؤكد للفعل: أبر بأبي ـــــــ.
</div>

Complete the sentence with an unrestricted noun that makes the sentence more emphatic:
I am kind with my father _____.

The parser is used to analyze the partial input sentence and determine its constituents as follows:

```
verbal_sentence(
 verb('أبرر',verb,male,singular,present,intrans,'بر'...),
        subject(noun('أبي', noun,male,singular,undef,...))
)
```

The morphological generator uses the infinitive verb 'بر' (kindness) of the main verb to synthesize the unrestricted object 'ابرار' (extremely kind). The learner's answer is also analyzed into a parse tree. The comparison between the corresponding representations will issue the appropriate feedback that describes the source of the error. The possible source of errors could be: different word (sense or category), incorrect morphological generation of a word (incorrect verb: tense, number, gender, ...; incorrect noun: number, gender, definition,...), incorrect syntactic generation of a word (s not in emphatic form, does not originates from the infinitive verb, ...).

5   Conclusions and Future Work

In this paper, we have discussed issues related to the development of error analysis and error handling in Arabic ICALL systems. NLP tools can be useful for ICALL and in particular for giving meaningful feedback to the learner. Learner-system communication in free natural language is computationally the most challenging and pedagogically the most valuable scenario in Arabic ICALL. The deep syntactic analysis of the learner's answer, whether correct or wrong, is compared against a system-generated answer. This enables feedback elaboration that helps learners to understand better and fill in the knowledge gaps.

The rule-based approach is used to give some freedom to the language learners in the way they phrase their answers. This also enables the exercise author to enter only one possible correct answer, thus saving much time compared to the previous pattern matching answer coding approach. The present system has been implemented using SICStus Prolog. The system is transportable and capable of running on an IBM PC which allows the learner to use it to learn Arabic language anywhere and anytime.

We plan to enrich the present system, e.g. make the system available on the Internet to serve remote learners worldwide (especially learners of Arabic as a second language), and extend the grammar coverage to include more advanced grammar levels.

References

1. Boytcheva, S.,Vitanova, I., Strupchanska, A., Yankova, M., Angelova, G.: Towards the assessment of free learner's utterances in CALL. In the Proceedings of InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems, Venice, 2004.
2. Cameron, K. (editor): call-media, design, & applications, Swets & Zeitlinge, 1999.
3. Gamper, J.,Knapp, J.: A Review of Intelligent CALL Systems. Computer Assisted Language Learning (CALL): An International Journal, 15(4), 329-342, Belgium: SWETS & ZEITLINGER publisher, 2002.
4. Holland, M. V., Maisano, R. , Alderks, C.: Parsers In Tutors: What Are They Good For? CALICO Journal, 11(1), 28-46, 1993.
5. Holland, M. V., Kaplan, J.D.: Sama, M.R., editors, intelligent language tutors: theory shaping technology. mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., 1995.
6. Lam, F.S., Pennington, M.C.: The Computer vs. the Pen: A Comparative Study of Word Processing in a Hong Kong Secondary Classroom. Computer Assisted Language Learning (CALL): An International Journal, 8(1), 1995, 75-92, Belgium: SWETS & ZEITLINGER publisher.
7. McEnery, T., Baker, J.P., Wilson, A.: A Statistical Analysis of Corpus Based Computer vs. Traditional Human Teaching Methods of Part of Speech Analysis. Computer Assisted Language Learning (CALL): An International Journal, 8(2), 1995, 259-274, Belgium: SWETS & ZEITLINGER publisher.
8. Shaalan, K.: An Intelligent Computer Assisted Language Learning System for Arabic Learners, Computer Assisted Language Learning: An International Journal, 18(1 & 2), PP. 81-108, Taylor & Francis Group Ltd., February 2005.
9. Swartz, M., Yazdani, M. (editors): intelligent tutoring systems for foreign language learning. chapter Introduction. Springer-Verlag, 1992.