Development of Computer Assisted Language Learning System for Arabic Using Natural Language Processing Techniques

Khaled Shaalan

Computer Science Dept., Faculty of Computers & Information,
Cairo Univ., 5 Tharwat St., Orman, Giza, 12613 Egypt

Email: shaalan@esic.claes.sci.eg

Abstract: This paper describes the development of a computer-assisted language learning (CALL) system for learning Arabic using natural language processing (NLP) techniques. This system can be used for learning Arabic by students at the primary schools. It provides grammar practice for learners of Arabic. The learners are encouraged to produce sentences freely in various situations and contexts and guided to recognize by themselves the erroneous or inappropriate functions of their misused expressions. In this system, we use NLP tools (including a morphological analyzer and syntax analyzer) and an error analyzer to give the adequate feedback to the learner. Furthermore, we propose the mechanism of correction by the learner which allows the learner to correct the typed sentence by herself/himself, and allows the learner to realize that what error she/he has made.

Keywords: Arabic natural language processing, CALL, handling ill-formed natural language input

تطوير نظام مساعد لتعليم اللغة العربية باستخدام تقنيات معالجة اللغات الطبيعية

هذا البحث يصف تطوير نظام مساعد لتعليم اللغة العربية باستخدام تقنيات معالجة اللغات الطبيعية. هـــذا النظـــام يمكن استخدامه لتعليم اللغة العربية لتلاميذ المرحلة الابتدائية. وهو يساعد على تعلم قواعد النحـــو لمـــتعلم اللغـــة العربية. وهو يشجع المتعلم على كتابة الجمل بحرية في مواضع وحالات مختلفة. كما يســـاعد المتعلمـــين علـــى اكتشاف الأخطاء بأنفسهم. في هذا النظام نستخدم أدوات معالجة اللغات الطبيعية مثل المحلل الصرفي و المحلـــل النحوي بالإضافة إلى محلل الأخطاء الذي يعطى المردود المناسب للمتعلم. علاوة على ذلك نقترح آلية لتصـــحيح الأخطاء والتي تتيح للمتعلم تصحيح الجملة المدخلة بنفسه كما تتيح له التحقق من نوع الخطأ.

1. Introduction

Computer-assisted language learning (CALL)[1] addresses the use of computers for language teaching and learning. The effectiveness of CALL system has been made obvious by many researchers (Lam et al., 1995; McEnery et al., 1995). Until quite recently, computer-assisted language learning was a topic of relevance mostly to those with a special interest in that area. Recently, though, computers have become so widespread in schools and homes and their uses have expanded so dramatically that the majority of language teachers must now begin to think about the implications of computers for language learning. Using computers provide a number of advantages for language learning (Warschauer, 1996):

- Repeated exposure to the same material is beneficial or even essential to learning.
- A computer is ideal for carrying out repeated drills, since the machine does not get bored with presenting the same material and since it can provide immediate non-judgmental feedback.
- A computer can present such material on an individualized basis, allowing students to proceed at their own pace and freeing up class time for other activities.
- The process of finding the right answer involves a fair amount of student choice, control, and interaction.
- Creates a learning environment, since listening is combined with seeing, just like in the real world.
- Multimedia and hypermedia technologies allow a variety of media (text, graphics, sound, animation, and video) to be accessed on a single machine. Hence, skills are easily integrated, since the variety of media makes it natural to combine reading, writing, speaking and listening in a single activity.
- Internet technology facilitates communications among the teacher and the language learners. It allows a teacher or student to share a message with a small group, the whole class, a partner class, or an international discussion list of hundreds or thousands of people.

The current CALL systems have the weakness that learner cannot key in target language sentences freely and cannot guide the learner to correct the most likely ill-formed input sentences. The learner just accepts the information which follows the instruction of curriculum which is pre-installed in the computer. Therefore, it cannot be said that these types of systems are totally interactive. Due to this reason, more and more researches on NLP techniques in CALL systems are needed. In CALL systems, the analysis of the typed sentence is a must in order to allow learners to phrase their own sentences freely without following any pre-fixed rules. Therefore, we used NLP tools in order to analyze typed sentence. Unfortunately, almost all of the current CALL systems that use NLP techniques can just analyze grammatically correct sentences. However, Arabic language learners who use the CALL system are most likely to key in ill-formed sentences. For this reason, we have used a mechanism that recognizes the structure of ill-formed input sentences, and then allows the learner to correct the typed sentence by herself/himself.

---

[1] CALL is also known as computer-assisted instruction (CAI), computer-aided instruction (CAI), or computer-aided language learning.

This paper describes a CALL system for Arabic using NLP techniques, called Arabic CALL, which can solve the weaknesses of the current CALL systems. The present system allows the learner to produce sentences freely in various situations and contexts and guided to recognize by themselves the erroneous or inappropriate functions of their misused expressions. In other words, it helps to make the learners make use of their own errors. 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 (including a morphological analyzer and syntax analyzer) and an error analyzer to give the adequate feedback to the learner. Furthermore, we propose the mechanism of correction by the learner which allows the learner to correct the typed sentence by herself/himself, and allows the learner to realize that what error she/he has made. Arabic CALL follows the curriculum of Arabic grammar at the Egyptian primary schools.

The rest of this paper is structured as follows. In Section 2, related work on Arabic CALL systems is given. In Section 3, we briefly describe our proposed Arabic CALL system. Sections 4 through 7 present the main components of the Arabic CALL system. In Section 8, we conclude the paper and give directions for future work.

2. Related Work

The linguistic computation of Arabic sentence is a difficult task (Othman et al., 2003). The difficulty comes from several sources: 1) the length of the sentence and the complex Arabic syntax, 2) The omission of diacritics (vowels) in written Arabic "altashkiil", 3) The free word order nature of Arabic sentence, and 4) The presence of an elliptic personal pronoun "alDamiir almustatir". For these reasons, there are very few researches involving Arabic CALL (Ditters et al. 1993).

Researches in Arabic CALL can be classified into two approaches: *Computer as a Tool* and *Computer as a Tutor.* In the *Computer as a Tool* approach, some computer programs can be used as a tool that do not necessarily provide any language material at all, but rather empower the learner (usually a native speaker) to use or understand language. In the *computer as a tutor* approach, the process of finding the right answer involves a fair amount of student choice, control, and interaction.

Computer as Tool

Hegazi et al. (1989) presented a way to represent Arabic syntax in Prolog as production rules. As the system can detect some errors concerning Arabic syntax, it can be used for an educational environment.

Abou_Ela (1994) developed an expert system for the Arabic Syntax Analyzer (ESASA), which can be used as a tool to assist Arabic linguists in building the Arabic grammar rules. The grammar is expressed using a declarative language called, Grammar Writing Language (GWL). This tool is aimed at building Arabic natural language applications including CALL.

Teachers and researchers trying to produce CALL software in languages that use the non-Latin alphabets have a long history of seeing their work disappear into a series of

nonsensical characters or black squares. This problem is particularly true of web-based material because there are so many unknown factors associated with the operating system of a distant user. Cushion et al. (2002) described how recent technological developments have provided the possibility of overcoming these technical problems in conjunction with the Java programming language and the Unicode character numbering system. This approach has enabled the conversion of previously developed software, which produced web-based CALL material in the Latin alphabet, to produce similar material in Arabic. The system does not rely on the restructuring of the host computer's operating system and enables teachers of Arabic to produce their own interactive CALL material in a form that will be easily accessible to their students

Shaalan (2003) developed an Arabic grammar checker, called Arabic GramCheck. Arabic GramCheck looks for common Arabic grammatical problems, describes the problem for you, and offers suggestions for improvement. This program is useful in pointing to problems believed typical of native speaker writing. Thus, she/he can avoid such problems in the future.

Computer as Tutor
Gheith et al. (1996) developed an Instructional Software for Teaching the Arabic Language (ISTAL) for grade one preparatory school. The system presents the curriculum as a simple concept associated with a set of generated sentences which highlight this concept. Then, the system generates an exercise to the student. The student's answer is automatically evaluated by comparing it to the system solution.

The Visual Interactive Syntax Learning for Arabic (ArabVISL) is an internet based interactive software for self-paced learning of Arabic grammar which is currently being developed at the University of Southern Denmark (Nielsen, 2001; Nielsen et al., 2003). It allows students of Arabic as a foreign language to analyze Arabic sentences by using Arabic script and Arabic grammatical terminology. The grammatical framework of ArabVISL as well as the technical solutions reflect a number of choices which had to be made in order to ensure a coherent and pedagogically suited product. This research focuses on these choices in an attempt to provide a better understanding of the problems encountered when developing CALL-materials for languages written in non-roman scripts.

The Interactive Language Learning project at London Guildhall University was asked to produce course material for the university's Arabic classes (Cushion et al., 2003). The authoring package was converted to the use of Arabic but, to avoid imposing models derived from European languages onto Arabic, a member of the design team decided to study the language from beginners level while collaborating with the teacher in producing the CALL package. This process has produced significant results and caused considerable re-evaluation of some preconceived concepts. This research highlights problems associated with learning a language with an unfamiliar alphabet and offers some suggestions as to the part CALL can play in overcoming them. It contains reflections on the way in which lessons thus learnt might be applied to languages that use the Latin script, including the possible use of CALL authoring as part of the learning process.

Limitations of the Current Language Learning Systems

In Egypt, some publishers of the on-the-shelf school textbooks provide students with either CD's or Web sites that contain vocabulary and grammar practice. However, most of these systems have some common limitations, which are:

1. They often resemble the traditional workbook exercises from which they were adapted.
2. From a pedagogical perspective, the definition of acceptable answers to exercises is highly constrained. For instance, in the linguistic analysis questions (إعراب), the learner can type his answer as follows: "مبتدأ مرفوع و علامة رفعه الضمة". Nevertheless, the system would consider this response as a wrong answer since it stores the answer of this question as: "مبتدأ مرفوع بالضمة".
3. Error feedback commonly does not address the source of an error. For instance, the system displays the correct answer without any explanation of the student's mistake. This makes the system's feedback a generic catchall response.
4. For vocabulary exercises, the student is referred to the corresponding page in the textbook, which displays the word in question in a word list. In addition to the pedagogical limitations, the student has to consult the textbook, which is an unnecessary inconvenience given the potential of the Web.
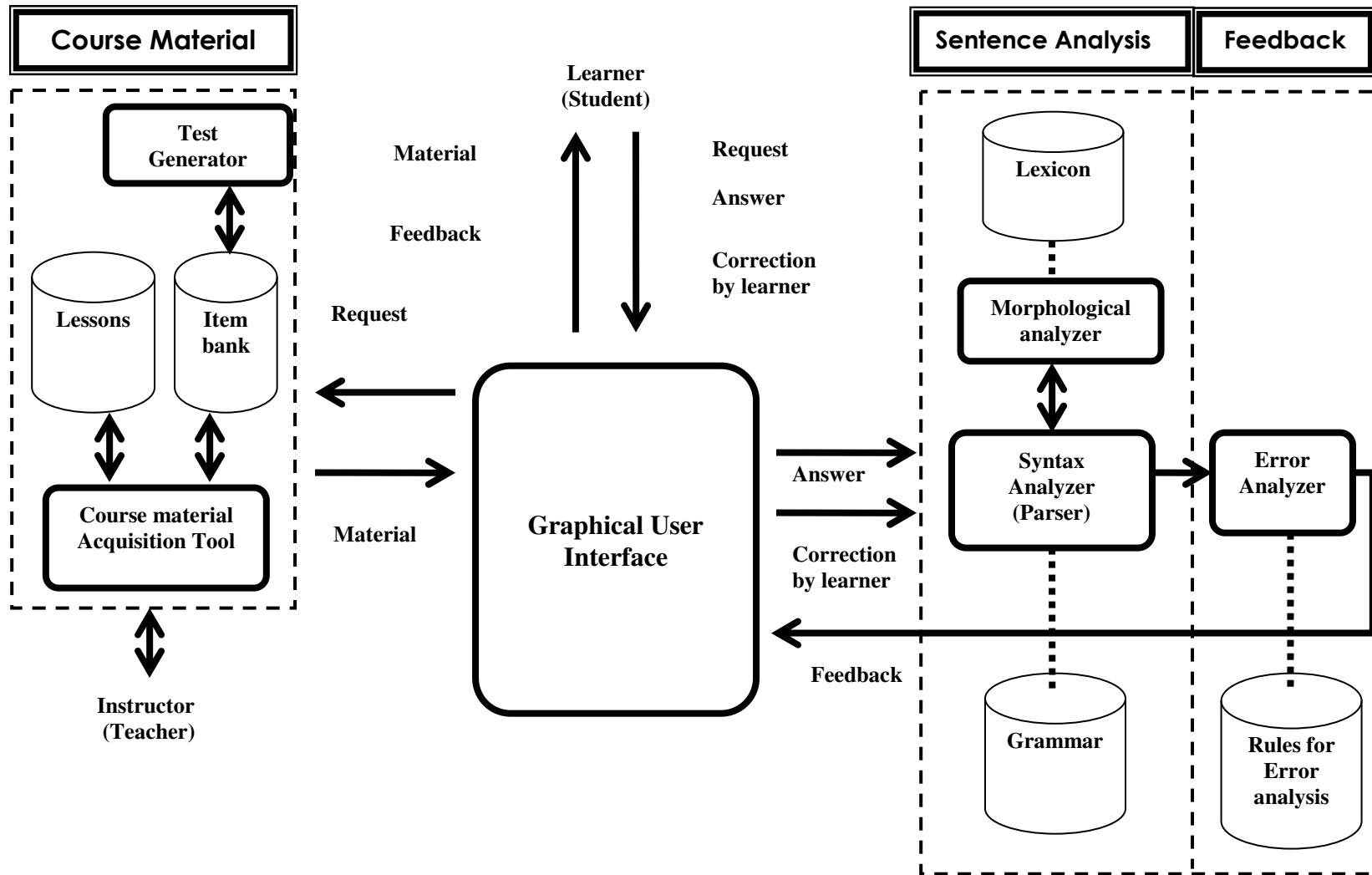
3. The Proposed System Architecture

Fig. 1 shows the proposed Arabic CALL system architecture. This system consists of the following subsystems: user interface, course material, sentence analysis, and feedback.

The user interface provides the means of communications between the learner and the Arabic CALL system. It is used to present multimedia lessons, provide the test items and allow the learner to take the test, and deliver the feedback to the learner. The user interface subsystem is implemented using Visual Basic and Flash.

The course material includes educational units, item bank, test generator, and acquisition tool. Each educational unit is a collection of Arabic grammar lessons that addresses a common topic. Item bank (question bank) is a database of test items. The test generator withdraws test items as needed to develop a test. The acquisition tool allows the instructor (teacher) to author and maintain lessons, and to create and maintain test items. The course material subsystem is implemented using Microsoft Access and Visual Basic.

The sentence analysis includes a morphological analyzer, syntax analyzer (parser), grammar rules, and lexicon. These tools are used to analyze inflected Arabic words, grammatically correct Arabic sentences and linguistic analysis of Arabic sentences. The sentence analysis subsystem is implemented using SICStus Prolog.

The feedback subsystem includes an error analyzer that is used to parse ill-formed learner input to give the adequate feedback to the learner. We augment the Arabic grammar with rules (buggy rules) which are capable of parsing ill-formed input and which apply if the grammatical rules fail. The feedback subsystem is implemented using SICStus Prolog.

Fig. 1 The Proposed Arabic CALL System Architecture

136

4. User Interface

The user interface provides the means of communications between the learner and the Arabic CALL system. It is used to present multimedia lessons (text, graphics, sound, and animation), provide the test items and allow the learner to take the test, and deliver the feedback to the learner.

There are two possible approaches for reacting to the learner's response: *linguistic analysis based approach* and *pattern matching based approach*. Applications developed with a NLP techniques will offer a much more reliable and efficient service to the learner. There are several reasons, which we believe should support this view: 1) NLP systems, usually conceived with a modular architecture, are much more easily extendible as linguistic research proceeds, and 2) it is not at all obvious that the performance of a CALL system, which does not perform a minimum of linguistic analysis, is really satisfactory.

On the contrary, a program with a simple letter-to-letter match is incapable of differentiating types of errors: not only is it, therefore, incapable of providing any valuable, evaluative feedback, but, in ignoring the source of the error when selecting another problem. In this study, we take the linguistic analysis based approach.

5. Course Material
5.1 Primary Level Lessons of Arabic Language
The educational units include Arabic grammar lessons for primary level. Specifically, they cover the following:

الأسماء– النعت – المثنى و الجمع – الأفعال - الضمائر - الجملة الاسمية –المبتدأ و الخبر – أنواع الخبر – الجملة الفعلية - العطف – الظرف – الحال – النداء – المفعول لأجله – المفعول المطلق –إن و أخواتها – أنواع خبر إن وأخواتها – كان و أخواتها – أنواع خبر كان و أخواتها – الحروف

Fig. 2 shows an example of a lesson explaining the unrestricted object "المفعول المطلق". It consists of an explanation of this grammar rule, an example, sound functionality, lesson test, and some navigation aids.

The lessons are stored in a database. The system includes some instructional templates to allow for quick generation of instructional material. The structure of lessons consists of two database relations, namely: lesson relation and example relation.

1. Lesson relation: This relation contains the following fields:
  ▪ *Lesson*. This field indicates the lesson to which this rule belongs. This field is the primary key.
  ▪ *Rule*. This field contains the wording of the grammar rule of the corresponding lesson.
  ▪ *Unit*. This field indicates the unit to which this lesson belongs.
  ▪ *Window id*. This field refers to the relevant lesson window.

2. Example relation: This relation contains the following fields:
- ▪ *Example id*. The primary key.
- ▪ *Lesson*. The foreign key.
- ▪ *Example*. This field contains the wording of the examples used in the lesson.

Fig. 2 A lesson

5.2 Item Bank

The item bank is a database of test items. This component is used to generate different types of test items each time the learner is allowed to take the test. The test generator selects the test items in a random order. According to the selection criterion made by the instructor, all the test items that match this criterion are collected. Then, we apply a random function to present the selected test items to the learner.

Fig. 3 A test item

Fig. 3 shows an example of a test item. It consists of a question header (identify the inchoative and enunciative, and the type of the enunciative in the following sentence), a test sentence (the brave soldiers make victory), a learner input area (text boxes and pull down menus), "Answer button" to allow the learners to see the model answer, "Check button" to check the learner's answers, hyperlinks to the relevant grammar lesson, and some navigation aids.

In Arabic CALL, there are two main types of test items for interaction with the learner: supply-type (short-answer/fill-in-the-space) or selection-type (matching, true/false, identify, or multiple-choice) interactive questions. The objective test method is used to assess the learner's knowledge or skills where each question has one (and only one) correct answer – and there is no ambiguity about what that correct answer should be. From the linguistic point of view, the type of questions used in our Arabic CALL system can be classified as follows:

1. Identify words according to certain morphological feature(s) or identifying constituents according to certain syntactic feature(s)
   - Examples:
     - Identify the category of each of the words in the following sentence "عين الاسم والفعل والحرف في الجملة الآتية:"
     - Identify the dual and regular plural in the following words " عين المثنى والجمع السالم في الكلمات الآتية: "
     - Identify the verb, subject, and object in the following sentence " عين الفعل والفاعل والمفعول به في الجملة الآتية"
     - Extract the adjective and the modified noun in the following sentence "استخرج النعت و المنعوت في الجملة الآتية:"
     - Identify the nominal and verbal sentence in the following sentences "عين الجملة الاسمية والجملة الفعلية في الجمل الآتية:"
2. Verb conjugation
   - Examples:
     - Give the correct present and imperative tense of the following verbs "أكتب الفعل المضارع والأمر للأفعال الآتية "
     - Present tense - Fill in the blank with the correctly conjugated form of the weak verb in parentheses " الفعل المضارع – أكمل باستخدام الفعل المعتل المناسب مما بين القوسين"
     - Fill in the blanks with the correct form of the verb " أكمل باستخدام الفعل المناسب"
3. Noun morphology
   - Examples:
     - What is the correct dual and regular plural of the following words? " ثنِّ واجمع سالما الكلمات الآتية:"
     - Complete the following sentences with the correct noun form in parentheses " أكمل الجمل الآتية باستخدام الاسم المناسب مما بين القوسين "
     - What pronoun would you use to talk to the following people? " ما هي الضمائر التي يمكن استخدامها مع الأشخاص التاليين"
     - Fill in the blanks with the correct form of the demonstrative noun " أكمل باستخدام اسم الإشارة المناسب"

4. Identify the grammatical relation
   - Examples:
     - Identify the type of the enunciative in the following sentence " عين نوع الخبر في الجملة الآتية"
     - Identify the negation or prohibition in the following sentence "عين النفي أو النهى في الجملة الآتية:"
     - Put the coordinating particle in the correct place in the following sentences " ضع حرف العطف في مكانه الصحيح من الجمل الآتية: "
5. Linguistic analysis of words between brackets or a sentence
   - Examples:
     - Give the linguistic analysis of the following sentence "أعرب الجملة الآتية"
     - Give the linguistic analysis of the words between brackets in the following sentence "أعرب ما بين القوسين في الجملة الآتية"
     - What's the difference in the linguistic analysis between the following? Give the reason "هل ترى فرقا إعرابيا في التالي؟ اذكر السبب"
     - Give the reason for the accusative case of the words between brackets "اذكر سبب نصب الكلمات التي بين القوسين:"
6. Transform the sentence category
   - Examples:
     - Change the following nominal sentence into verbal sentence " الجملة الآتية اسمية أجعلها فعلية"
     - Change the following sentences as shown in the example " حول كما في المثال "
     - Forbid your colleague from the following using (lā of prohibition) " اِنه زميلك عما يأتي مستخدمًا أداة النهى (لا):"
     - Advise your colleague of the following using imperative verb " انصح زميلك بما يلي مستخدمًا فعل أمر:"
     - Change the sentence to the plural following the example " اجعل الجملة جمعا باتباع المثال التالي"
7. Agreement
   - Examples:
     - Rewrite the following sentences, changing the demonstrative noun from the singular form into the plural form, and change what is necessary to make it a grammatically correct sentence " اجعل الإشارة للجمع وغير ما يلزم"
     - Is the agreement between the adjective and the modified noun in the following sentence correct or incorrect? Give the reason " هل التطابق بين النعت والمنعوت في الجمل الآتية صحيح؟ اذكر السبب"
     - Subject-verb agreement. Use the correct present-tense form of the verb in parentheses. " تطابق الفعل والفاعل – استخدم الفعل المضارع المناسب من الأفعال التي بين القوسين"
8. Review test
   - Examples:
     - Complete the following passage by selecting the correct verb/adjective for the context "أكمل القطعة التالية باختيار الفعل أو النعت المناسب للسياق"
     - Put the sentence into the correct order "رتب الجملة ترتيبا صحيحا"

- o Match column (a) with column (b) to have a meaningful sentence. Use the boxes on the right to write your choices " (ب) بالعمود (أ) صل العمود لتحصل على جملة مفيدة. استخدم الصناديق التي في الجانب الأيمن للإجابة "
- o Are the following sentences grammatically correct? " هل الجمل التالية صحيحة إعرابيا"

The structure of the item bank consists of three database relations, namely: question title relation, question content relation, and answer relation.

1. **Question title relation:** This relation contains the following fields:
   - *Question id*. The primary key.
   - *Question type*. This field indicates the question type: short-answer, matching, true/false, identify, or multiple choice.
   - *Question title*. This field indicates the question header.
   - *Lesson*. This field indicates the lesson to which this question belongs.
   - *Window id*. This field refers to the relevant question window.
2. **Question contents relation:** This relation contains the following fields:
   - *Sentence id*. The primary key.
   - *Question id*. The foreign key.
   - *Sentence*. This field contains the wording of the sentence used in the question.
   - *Window id*. This field refers to the relevant question window.
3. **Answer relation:** This relation contains the following attributes:
   - *Answer id*. The primary key.
   - *Sentence id*. The foreign key.
   - *Answer*. This field contains the model answer of the question.

6. Sentence Analysis
6.1 Morphological Analysis
The Arabic language is based on the Semitic root-and-pattern scheme of forming word stems, as well as the concatenation of stem and affixes. We need a sophisticated morphological analyzer that is capable of transforming the inflected Arabic word into its origin. To achieve this function we developed a morphological analyzer for inflected Arabic words (cf. Rafea et al., 1993). The morphological analyzer analyzes the inflected Arabic word to extract the stem and its features. An augmented transition network (ATN) (Woods, 1970) technique was successfully used to represent the context-sensitive knowledge about the relation between a stem and inflectional additions. The ATN consists of arcs. Each of which is a link from a departure node to a destination node, called states, see Fig. 4. An exhaustive-search to traverse the ATN generates all the possible interpretations of an inflected Arabic word. The morphological analyzer is implemented in Prolog and integrated with the DCG parser.
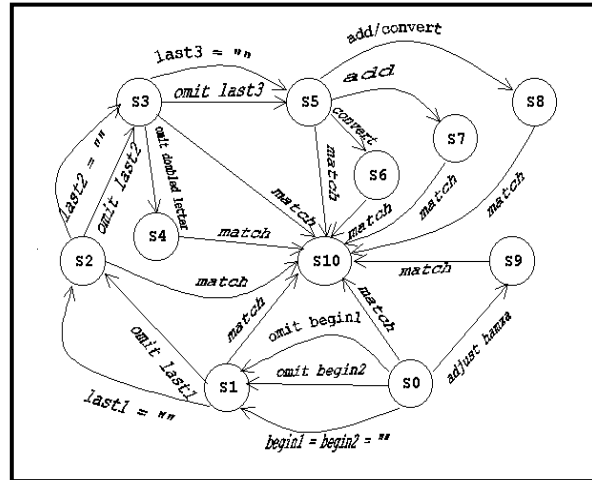
Fig. 4 : ATN representing the relation between the affixes and stem
of an inflected Arabic word.

An Arabic monolingual lexicon is also needed to successfully implement the morphological analyzer. The lexicon is designed to reflect the word categories in Arabic — each with a different set of features.

The morphological analysis in Arabic CALL system analyzes the learner's answer in response to a generated question, such as "Fill-in-the space". This answer should meet certain morphological rules. These rules are used to guide the analysis of the learner's answer. This method has the following advantages: minimizing the ambiguity, facilitating the generation of the adequate feedback in case of ill-formed input, and speeding up the analysis phase.

6.2 The Grammar Formalism
The grammar for Arabic contains the grammar knowledge required to analyze a grammatically correct sentence. The grammar is being developed especially for learning Arabic. Currently, it concerns the Arabic grammar at the primary level. We adopted general solutions as much as possible, as this increases the chances that the grammar can be used in other domains as well. Thus, in designing the grammar we seek a balance between short-term goals (a grammar which covers sentences typical for learning Arabic and is reasonably robust and efficient) and long-term goals (a grammar which covers the major constructions of Arabic in a general way).

From a linguistic perspective, the current grammar can be characterized as Definite Clause Grammar (DCG) formalism (Pereira et al., 1986). The choice for DCG is motivated by the fact that this formalism provides a balance between computational efficiency and linguistic expressiveness. The central formal operation in DCG is unification of feature-structures. Table 1 describes the features used in the current grammar along with their possible values.

Table 1 Features and their Values for the Arabic Grammar

Feature	Possible values	Comments
Gender	masculine, feminine	
Number	Singular, dual, plural	
Definiteness	Definite, indefinite	Applied only to nouns
Special noun	Yes, no	determine whether or not the noun is Inna and its sisters (إن وأخواتها)
Pattern	Form of pattern (wazen)	
End case	Accusative, nominative, genitive	Iarab الأعراب
Transitivity	Transitive, intransitive	Applied only to verbs
Special verb	Yes, no	determine whether or not the verb is Kan and its sisters (كان و أخواتها)
Affix	Affixes of the inflected word	
Current category	Category of the grammatical symbol being parsed	
Next category	Category the grammatical symbol that follows the current symbol	
Noun as adjective	Yes, no	Determine whether or not the noun can be used as an adjective.
Noun as annexation	Yes, no	Determine whether or not the noun can be annexed (إضافة)
Verb tense	Past, present, imperative	
Single word	Single form of the broken plural	
Infinitive	verb Infinitive form	
Person	First, second, third	Applied only to pronouns
Noun refers to time or place	Time, place	Determine whether the accusative (ظرف) is related to time, related to place, or both.

Grammar rules for grammatically correct sentence
In the following, we show an extraction of DCG rules for parsing grammatically correct Arabic verbal sentence.

verbal_sentence(verbal_sentence (SVS)) --> simple_verbal_sentence(SVS).
verbal_sentence(verbal_sentence (PVS)) --> prefixed_verbal_sentence(PVS).
verbal_sentence(verbal_sentence (SVS)) --> special_verbal_sentence(SVS).

simple_verbal_sentence((simple_verbal_sentence(V,S,O,UO)) -->
    verb(V,Gender, transitive, Infinitive, no),
    subject(S,Gender),
    object(O),
    unrestricted_object(UO,Infinitive).
simple_verbal_sentence((simple_verbal_sentence(V,S,Comp)) -->
    verb(V,Gender, intransitive, _, no),
    subject(S,Gender),
    complement(Comp).

prefixed_verbal_sentence(prefixed_verbal_sentence(Neg_art, SVS)) -->
    negative_paticle(Neg_art),
    simple_verbal_sentence(SVS).

```
special_verbal_sentence(special_verbal_sentence(V,NS)) -->
        verb(V, _,_,_,yes),
        nominal_sentence(NS).
```

The simple verbal sentence has two forms:
- Form1: <verb> + <subject> + <object> + <unrestricted_object>
  - Constrains:
    1. Verb and Subject agrees in gender.
    2. Verb is transitive.
    3. Verb isn't a special verb such as Kan and its sisters "كان و أخواتها".
    4. The unrestricted object "مفعول مطلق" comes after the object.
  - Example:
    > أعاتب صديقي عتابا رقيقا.
- Form2: <verb> + <subject> + < complement >
  - Constrains:
    1. All the above constrains except that the verb is intransitive.
  - Example:
    > أقف لمعلمى احتراما.

The prefixed verbal sentence is a simple verbal sentence that begins with the negative particle like 'لا'. The special verbal sentence is the Kan and its sisters sentence.

Grammar rules for linguistic analysis
We have also developed a grammar that helps us in parsing the learner's answer in response to a question for the linguistic analysis of a given Arabic sentence. The parser takes the learner's answer and converts into a quadruple abstract representation form:

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

For example, the linguistic analysis of the word between braces in the sentence
- استمتعت بجو الريف (استمتاعاً)

is

> مفعول مطلق منصوب بالفتحة لأنه مفرد.

The learner may write his answer as either
  - مفعول مطلق منصوب وعلامة نصبه الفتحة لأنه مفرد
  - مفعول مطلق منصوب وعلامة نصبه الفتحة وهو مفرد
  - مفعول مطلق منصوب بالفتحة لأنه مفرد.

This will be converted by the parser into the abstract representation
> مفعول مطلق + منصوب + فتحة + مفرد

This abstract representation facilitates the matching between the learner answer and the correct answer generated by the system. To show how the linguistic analysis is generated, consider the following example.

أعرب ما بين القوسين في الجملة الآتية:
- ألف العقاد (كتبًا كثيرة).

The following DCG rule describes the linguistic analysis of the words between brackets.

```
object(Words, Rest, Analysis) -->[X],[Y],
  {get_analysis( X, Gender, Num, Definite, _, Words, Rest1, Analysis1, End_case, 'مفعول به'),
   (var(End_case) -> rule('مفعول به', End_case);true),
   get_analysis( Y, Gender, Num, Definite, Can_Be_Adj, Rest1, Rest, Analysis2, End_case, 'نعت'),
   append(Analysis1, Analysis2, Analysis)
  }.
```

This rule takes the list of words to be analyzed as input (the words between brackets) and produces as output both the rest of this list and the linguistic analysis of these words. This is done by sending these words to get_analysis/10 one after another. The definition of get_analysis/10 is as follows:

```
get_analysis(Accepted_word, Gender, Num, Definite Can_be_adj, [Word| Rest],Rest_words,
        Analysis, End_case, Location):-
     morph(Accepted_word,lex(_,noun,Gender,Num,Definite,Can_be_adj,_,_),_),
     (Word == Accepted_word ->
       get_Irab(Location,Num,End_case,Word_analysis),
       Rest_words = Rest, Analysis = [Word_analysis])
     ; Rest_words = [Word|Rest], Analysis=[]).
get_analysis(Accepted_word,_Gender,Num,Definite,Can_be_adj,[],[],[],_,_):-
   morph(Accepeted_word,lex(_,noun,Gender,Num,Definite,Can_be_adj,_,_),_).
```

After the word is morphologically analyzed, its features are sent to get_Parse/4 to generate the quadruple abstract representation form.

```
get_Parse(Location, Num, End_case, Word_analysis):-
  rule(Location,End_case),
  rule(Nums, End_case, Analytic_sign),
  member(Reason, Nums),
  Word_analysis = [Location, End_case, Analytic_sign, Reason].
```

get_Parse/4 uses some facts about the location of Arabic words (الموقع الإعرابي) and its analysis (إعراب). An example of these facts is as follows:

```
rule('مبتدأ','مرفوع').
rule('نعت',_).
rule['مفرد' , 'جمع تكثير' ,'جمع مؤنث سالم'],'الضمة','مرفوع').
rule(['مثنى','الألف','مرفوع').
rule['جمع مذكر سالم'],'الواو','مرفوع').
```

6.3 Parsing
Logic programming plays an essential role in natural language analysis process because it attempts to use logic to express grammar rules and to formalize the process of parsing (Gazdar et al., 1990). A grammar specified this way is known as *logic grammar* since it

145

represents rules as Horn clauses (Dougherty, 1994). Logic grammars can be conveniently implemented in Prolog. Prolog-based grammars can be quite efficient in practice (Allen, 1995). Prolog interpretation algorithm uses exactly the same search strategy as the depth-first top-down parsing algorithm, so all that is needed is a way to reformulate grammar rules as clauses in Prolog. Definite clause grammars (DCGs) notation was developed as a result of research in natural language parsing and understanding (Pereira et al., 1986). DCGs allow one to write grammar rules directly in Prolog, producing a simple recursive decent parser. During the construction of the Arabic parser, feature-structures are translated into Prolog terms. Because of this translation step, parsing can make use of Prolog's built-in term-unification, instead of the more expensive feature-unification. Prologs that conform to Edinburgh standard have DCGs as a part of their implementations. In the current system, grammar rules of Arabic are written in DCG formalism, which are translated into executable code in SICStus Prolog.

7. The Feedback system
Feedback is the computer's response to answers made by learners/students. Feedback gives students a feel for how well they are progressing through a lesson, thereby increasing their confidence levels. It also reinforces the subject matter.

The feedback system compares the learner's answer produced from the sentence analysis subsystems with the correct answer that was generated by the system. If there is a match, a positive message will be sent to learner. Otherwise, an adequate feedback message will be sent to the learner, see Fig. 5.

Fig. 5 A feedback message indicating incorrect answer

The learner can either read the feedback message and correct the typed sentence instantly, or she/he can restudy the related grammar items and then correct the sentence by herself/himself. In the following subsections, we show how the system catches the learner's errors and how it handles the ill-formed natural language input.

7.1 Rules for Error Analysis
In our implementation, we augment the Arabic grammar with rules (buggy rules) which are capable of parsing ill-formed input and which apply if the grammatical rules fail. Buggy rules provide a distinct rule for every ill-formed case.

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

أكمل بمفعول مطلق مناسب:
- أبر والدي _____.

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 verb infinitive. 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 verb infinitive but is defined. Issue a message describing that the unrestricted object should be undefined.
- A word that is a noun, originates from the verb infinitive and is undefined, but needs the end case "Alef Tanween". Issue a message describing that a missing end case of the unrestricted object.
- A Correct answer. Issue a positive message.

From the analysis of the learner's answer, we augmented the grammar of the unrestricted object by buggy rules that handle each ill-formed case as follows.

```
unrestricted_object(UO,Infinitive) --> [Word],
  {morph(Word, lex(Stem,Category, _, _, Definiteness, _, _, _),_),
   check_correctness(Infinitive,Feedback,Word,[Stem,Category,Definiteness]),
  (Feedback == 'إجابة صحيحة' -> UO = unrestricted_object(Word);
   UO = incorrect_unrestricted_object(Word,Feedack) }.


check_correctness(_, Feedback, Word, [_,Category,_]):-
        Category \= noun,!,
        error_flagging(not(noun),"المفعول المطلق", Feedback,[Word]).
check_correctness(Infinitive, Feedback, Word,[Stem, noun,_]):-
        Stem \= Infinitive, !,
        error_flagging(not(infinitive), "المفعول المطلق", Feedback,[Word]).
check_correctness(Infinitive, Feedback, Word,[Infinitive, noun, defined]):-
      error_flagging(not(undefined_noun), "المفعول المطلق", Feedback, [Word]),!.
check_correctness(Infinitive, Feedback, Word,[Infinitive, noun, undefined]):-
        name(Word, Str),
        name(Infinitive,Str1),
        (need_alaf_tanween(Infinitive) ->
                (append(Str1,"ا",Str) -> Feedback = 'إجابة صحيحة'
                ; error_flagging(need_alaf_tanween, "المفعول المطلق", Feedback, [Word])
                )
        ; Feedback = 'إجابة صحيحة'
        ).
```

7.2 Error Handling Mechanism
Learner's answers which have special handling mechanisms in case of the ill-formed learner input are: linguistic analysis, classification into categories, sentence transformation, and completing a sentence. They are discussed in the following subsections.

7.2.1 Handling of linguistic analysis

Linguistic analysis can be 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 algorithm for handling linguistic analysis.

Input:	o A given sentence in the question words
	o A given sequence of words between brackets, if any.
	o A learner's answer: linguistic analysis of the given sentence
Step1:	Convert the learner's answer into the abstract representation form
Setp2:	Parse the given sentence (or the sequence of words between brackets) and generate its linguistic analysis in the abstract representation form
Step3:	Compare the learner's answer with the generated answer.
Output:	IF both answers match
	THEN generate a positive message
	OTHERWISE generate an appropriate error message

Example:

What's the difference in linguistic analysis between the words in brackets?

هل ترى فرقا إعرابيا فيما بين القوسين؟
- استمتعت بجو الريف (استمتاعاً).
- أذهب إلى الريف (استمتاعاً) بجوه.

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer	Feedback
1. Correct answer.	استمتاعاً (١): مفعول مطلق منصوب بالفتحة لأنه مفرد. استمتاعاً (٢): مفعول لأجله منصوب بالفتحة لأنه مفرد.	إجابة صحيحة.
2. Error in analytic location	استمتاعاً (٢): مفعول به منصوب بالفتحة لأنه مفرد.	يوجد خطأ في الموقع الإعرابي.
3. Error in end case	استمتاعاً (١): مفعول مطلق مرفوع بالضمة.	يوجد خطأ في الإعراب.
4. Error in analytic sign	استمتاعاً (٢): مفعول لأجله منصوب بالألف.	يوجد خطأ في علامة الإعراب.
5. Error in reason	استمتاعاً (٢): مفعول لأجله منصوب بالفتحة لأنه جمع تكسير.	يوجد خطأ في السبب.
6. Correct answer but incomplete	استمتاعاً (١): مفعول مطلق منصوب.	إجابة ناقصة. العلامة الإعرابية ناقصة ، السبب ناقص.

7.2.2 Handling classification into categories

Classification into categories can be either for identifying morphological categories or for identifying syntactic constituents (possibly, a complete sentence). The following description outlines the algorithms for handling of classification into categories.

I. Classification into morphological categories

Input:	o A given sentence in the question words
	o A learner's answer: Morphological classification into N lists
Step1:	Morphologically analyze the words in the given sentence and determine the features of words.
Step2:	Generate an N lists, a classification of the words according to the questions words.
Step3:	Compare the learner answer with the generated answer.
Output:	IF both answers match
	THEN generate a positive message
	OTHERWISE generate an appropriate error message

Example:
Identify the category of each of the words in following sentence

عين الاسم والفعل والحرف في الجملة الآتية:

o   وقف التلاميذ احتراماً للمعلم

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer			Feedback
	Verb	**Noun**	**Particle**	
1. Correct answer	[وقف]	[التلاميذ، احتراماً، المعلم]	[اللام]	إجابة صحيحة.
2. Missing words.	[ ]	[التلاميذ، احتراماً، المعلم]	[اللام]	إجابة الأفعال ناقصة.
3. Incorrect category	[وقف، المعلم]	[التلاميذ، احتراماً]	[اللام]	كلمة "المعلم" ليست فعل.

II. Classification into syntactic constituents

Input:	o A given sentence in the question words
	o A learner's answer: Syntactic classification into N lists
Step1:	Parse the given sentence and determine the parse tree.
Step2:	Generate an N lists, a classification of the sentence's constituents according to the words of the questions.
Step3:	Compare the learner answer with the generated answer.
Output:	IF both answers match
	THEN generate a positive message
	OTHERWISE generate an appropriate error message

Example:

Identify the inchoative and, enunciative and provide its type in the following sentence عين

المبتدأ والخبر مبينا نوعه في الجملة الآتية:

o الجنود الشجعان ينتصرون

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer			Feedback
	Inchoative	**Enunciative**	**Enunciative type**	
Correct answer	الجنود	ينتصرون	جملة فعلية	إجابة صحيحة
Error in analytic location (enunciative type)	الجنود	ينتصرون	جملة اسمية	خطأ في نوع الخبر: خبر "الجملة الاسمية " يجب أن يبدأ باسم، وكلمة "ينتصرون" فعل وليست اسم.
	الجنود	ينتصرون	مفرد	خطأ في نوع الخبر: الخبر"المفرد" هو ما ليس جمله ولا شبه جملة. و يتكون من كلمة واحدة و تكون اسم، و كلمة "ينتصرون" فعل و ليست اسم.
	الجنود	ينتصرون	شبه جملة	خطأ في نوع الخبر: الخبر" شبه الجملة" يتكون من جار و مجرور أو ظرف (زمان أو مكان)، وكلمة " ينتصرون" ليست أي منهما.
Error in inchoative	الشجعان	ينتصرون	جملة فعلية	كلمة "الشجعان" نعت و ليست مبتدأ.
Error in enunciative	الجنود	الشجعان	جملة فعلية	كلمة "الشجعان" نعت و ليست خبر.

7.2.3 Handling transformation of a sentence

Transformation requires the learner to change/rewrite the form of a sentence. The following description outlines the algorithm for handling of transformation of a sentence.

Input:	o A given sentence in the question words
	o A learner's answer: transformed sentence
Step1:	Parse the given sentence and determine its words
Step2:	Parse the transformed sentence and determine the parse tree.
Step3:	Compare words of the given sentence with words in the transformed sentence
Output:	IF a word in the given sentence is either missed from or added to the transformed sentence OR
	the transformed sentence is grammatically incorrect OR
	the transformed sentence is not in the correct form
	THEN generate an appropriate error message
	OTHERWISE generate a positive message

Example:
Change the following nominal sentence into verbal sentence

الجملة الآتية اسمية أجعلها فعلية:
- المعلم يشرح الدرس

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer	Feedback
1. Correct answer	يشرح المعلم الدرس	إجابة صحيحة.
2. Extra words	يشرح المعلم الدرس في الفصل	يوجد بالإجابة كلمات زائدة: "في"، "الفصل".
3. Missing words	يشرح الدرس	يوجد بالإجابة كلمة ناقصة: "المعلم"
4. Incorrect transformation	يشرب محمد اللبن	جملة خاطئة
5. Grammatically incorrect verbal sentence.	المعلم يشرح الدرس	خطأ في التحويل: ليست جملة فعلية صحيحة
6. Grammatically incorrect verbal sentence.	المعلم الدرس يشرح	خطأ في التحويل: ليست جملة فعلية صحيحة
7. Incorrect transformation of the tense of the verb.	شرح المعلم الدرس	جملة فعلية صحيحة، و لكن يوجد خطأ في تصريف زمن الفعل: "شرح"
8. Incorrect number of the verb.	يشرحوا المعلم الدرس	جملة فعلية صحيحة، و لكن يوجد خطأ في تصريف عدد الفعل: "يشرحوا"
9. Incorrect transformation of the gender of the verb.	تشرح المعلم الدرس	جملة فعلية صحيحة، و لكن يوجد خطأ في تصريف جنس الفعل: "تشرح"
10. Incorrect transformation of the definition article of the object.	يشرح المعلم درس	جملة فعلية صحيحة، و لكن يوجد خطأ في تعريف المفعول به: "درس"
11. Incorrect transformation of the number of a noun.	يشرح المعلمون الدرس	جملة فعلية صحيحة، و لكن يوجد خطأ في تصريف عدد الاسم: "المعلمون"
12. Incorrect transformation of the gender of a noun.	يشرح المعلمة الدرس	جملة فعلية صحيحة، و لكن يوجد خطأ في تصريف جنس الاسم: "المعلمة"

7.2.4 Handling of Fill-in-the-space

Fill-in-the-space can be for generation of isolated words with particular form, or generation of words to complete a sentence that achieves features agreement among its constituents. The following description outlines the algorithms for handling fill-in-the-space.

I. Rewriting of isolated words with particular morphological form

Input:	o A given words in the question o A learner's answer: Morphologically generated forms of the given words
Step1:	Morphologically analyze the given words and determine their features.
Step2:	Check these features with the requirements in the question words
Output:	IF values of features match THEN generate a positive message OTHERWISE generate an appropriate error message

Example:
What is the correct dual and regular plural of the following words?

ثنِّ واجمع سالما الكلمات الآتية:
- المهندس
- الثمرة
- صحراء

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer		Feedback
	Dual	**Plural**	
1. Correct answer	(١) المهندسان – المهندسين (٢) الثمرتان – الثمرتين (٣) صحراوان - صحراوين	(١) المهندسون – المهندسين (٢) الثمرات (٣) صحراوات	إجابة صحيحة
2. Incorrect word	(١) الطالبان		كلمة " الطالبان" زائدة
3. Incorrect number of the noun		(٢) الثمرتان	إجابــة خاطئـة كلمـة " الثمرتان " ليست جمع
4. Incorrect definition & gender of the noun	(١) المهندستان		إجابة خاطئة ، هذه الكلمة تختلف عن الكلمة الأصلية في التعريف والجنس
5. Incorrect number of the noun	(٣) صـــــحراءان - صحراءين	(٣) صحراءات	خطـأ فـي تثنيـة الاسـم: صحراء خطـأ فـي جمـع الاسـم: صحراء

II. Fill-in-the-space to achieve agreement among sentence constituents

Input:	○ An incomplete sentence ○ A learner's answer: word(s) that completes the sentence
Step1:	Analyze the sentence and determine the features of the given words.
Step2:	Check the correctness of these features with the requirements in the question words
Output:	IF values of features match THEN generate a positive message OTHERWISE generate an appropriate error message

Example:

أكمل بمفعول مطلق مؤكد للفعل:
○ أبر والدي ـــــــ .

The following table shows the possible learner's answer along with the corresponding feedback

Source of Error	Learner's Answer	Feedback
1. Correct answer	برا	إجابة صحيحة
2. Incorrect word category.	يطيع	إجابة خاطئة. هذه الكلمة ليست اسم
3. Incorrect word form.	اجتهادا	إجابة خاطئة. هذه الكلمة ليست من أصل الفعل
4. Incorrect definition with the article.	البر	إجابة خاطئة. لكي يكون المفعول المطلق مؤكد للفعل لابد أن يكون نكرة. هذه الكلمة معرفة
5. Missing end case.	بر	إجابة خاطئة. هذه الكلمة تحتاج إلى ألف التنوين

8. Conclusions

In this paper, we describe the development of a NLP based CALL system for learning Arabic language. The system is implemented using SICStus Prolog, Visual Basic, Flash and Microsoft Access. This system allows the learner to key in Arabic sentences freely. We used NLP tools that are capable of parsing both grammatically correct Arabic sentences and linguistic analysis of the Arabic sentences. Moreover, we developed the mechanisms that are capable of detecting the structure of ill-formed input, and then allow the learner to correct the typed sentence by herself/himself. The system is transportable and capable of running on IBM PC's which allows the learner to use it to learn the Arabic language anywhere and anytime. This approach can thus be considered as a new means in language learning (especially Arabic language in our case) in the future.

We plan to enrich the present system, e.g. add a student or course management facility that allows us to have full performance and record-keeping features, add more

multimedia instructional material to allow learners fully comprehend what they learn in natural settings, making 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. Abou_Ela M. (1994). *Knowledge-based Techniques in an Arabic Syntax Analysis Environment*, ISSR, Cairo Univ., Ph. D. Thesis.
2. Allen, J. (1995). *Natural Language Understanding*, second edition, The Benjamin/Cummings Publishing Company.
3. Cushion S. and Hémard D. (2002). Applying New Technological Developments to CALL for Arabic, *Computer Assisted Language Learning (CALL): An International Journal*, Vol. 15(5):501-508, SWETS & ZEITLINER publisher.
4. Cushion S. and Hémard D. (2003). Designing a CALL Package for Arabic While Learning the Language Ab Initio, *Computer Assisted Language Learning (CALL): An International Journal*, Vol. 16(2-3): 259-266, SWETS & ZEITLINER publisher.
5. Ditters E., Oostdijk N. and Cameron K. (1993). *Processing Arabic: Computer Assisted Language Learning*, Intellect Books, Oxford, England.
6. Dougherty R. (1994). *Natural Language Computing: An English Generative Grammar in Prolog*, Lawrence Erlbaum Associates, Inc., NJ.
7. Gazdar G. and Mellish C. (1990). *Natural Language Processing in Prolog: An Introduction to Computational Linguistics*, Addison-Wesley.
8. Gheith M., Dawa I., Afifty M. (1996). ISTAL: Instructional Software for Teaching the Arabic Language, *In the Proceeding of the 1ˢᵗ International Conference on Computer and Advanced Technology in Education*, pp. 41-66, Cairo, Egypt.
9. Hegazi, N., Ali, G., Abed E.M., Hamada S. (1989). Arabic Expert System for Syntax Education, *In the Proceedings of the Second Conference on Arabic Computational Linguistics*, pp. 596-614 Nov., Kuwait.
10. Lam, F.S., and Pennington, M.C. (1995). 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), 75-92.
11. McEnery, T., Baker, J.P., and Wilson, A. (1995). 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): 259-274.
12. Nielsen H. (2001). ArabVISL: Arabic grammar on the internet, *In the proceeding of the 5ᵗʰ Nordic conference on Middle Eastern Studies, Workshop on Teaching Middle Eastern Languages - past and presen,* Lund.
13. Nielsen H. and Carlsen M. (2003). Interactive Arabic Grammar on the Internet: Problems and Solutions, *Computer Assisted Language Learning (CALL): An International Journal*, Vol. 16(1):95-112, SWETS & ZEITLINER publisher.

14. Othman E., Shaalan K., and Rafea A. (2003) A Chart Parser for Analyzing Modern Standard Arabic Sentence, *In proceedings of the MT Summit IX Workshop on Machine Translation for Semitic Languages: Issues and Approaches*, New Orleans, Louisiana, U.S.A.

15. Pereira F., Shieber C., and David H. (1986). Definite Clause Grammars for language analysis–A Survey of the Formalism and a Comparison with Augmented Transition Networks, *In Readings in Natural Language Processing, Grosz, B., Jones K., Webber B. (Eds.)*, pp. 24-101, Morgan Kuffmann.

16. Rafea A., Shaalan K. (1993). Lexical Analysis of Inflected Arabic words using Exhaustive Search of an Augmented Transition Network, *Software Practice and Experience*, 23(6):567-588, John Wiley & sons, U.K., June.

17. Shaalan K. (2003). Arabic GramCheck: A Grammar Checker for Arabic, *Egyptian Informatics Journal*, Faculty of Computers & Information, Cairo University, 2002, 4(1):94-111, June.

18. Warschauer, M. (1996). Computer-assisted language learning: An introduction. In Fotos S. (Ed.), Multimedia language teaching, pp. 3-20, Tokyo: Logos International.

19. Woods W. (1970). Transition network grammar for natural language analysis, *comm. ACM*, Vol. 10, pp. 591-66.