**Technology: Intelligent Computer Assisted Language Learning System for Arabic Learners**

1. **Brief Overview**

Computer-assisted language learning (CALL)[[1]](#footnote-2) addresses the use of computers for language teaching and learning. CALL emerged in the early days of computers. Since the early 1960's, CALL software was designed and implemented. The effectiveness of CALL systems has been made obvious by many researchers (Lam & Pennington, 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 provides 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.
* The computer can create a realistic learning environment, since listening can be combined with seeing, just as 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.
* Incorporating NLP techniques provide learners with more flexible—indeed, more ‘intelligent’—feedback and guidance in their language learning process.

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 (Swartz & Yazdani M., 1992; Holland et al., 1995; Cameron, 1999; Gamper & Knapp, 2002) .

1. **State of the Art (For Latin Languages)**
   1. Technology and Future Trends

?????

* 1. Applications and Reported Performance

????

1. **State of the Art (For Arabic Language)**
   1. Technology and Future Trends

Arabic is a highly derivational language that makes it a challenge to students and second language learners. In Arab world, where there is a growing demand in using computers for teaching and learning, some publishers of off-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 (إعراب) questions, the learner can type his answer as follows: "مبتدأ مرفوع وعلامة رفعة الضمة"   
   (inchoative is in nominal case and the diacritic sign is dam-mah)  
   Nevertheless, the system would consider this response as a wrong answer since it stores the answer of this question as: "مبتدأ مرفوع بالضمة" (inchoative is in nominal case and the diacritic sign is dam-mah).
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.
   1. Current and Envisioned Applications and Market Priorities

In Arabic ICALL, there are two main types of test items for interaction with the learner: *selection-type* that tends to elicit answers easily classified as right or wrong and *supply-type* requiring the learners to write a few words. Current systems, especially developed by Faculty of Computers and Information, Cairo University, focuses on objective test method 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. These systems guide learners to recognize by themselves the erroneous or inappropriate functions of their misused expressions. In other words, it helps learners to make use of their errors. It doesn't give them the correct answer directly but it enables them to try over and over again. In this system, NLP tools (including morphological analyzer and syntax analyzer) and error analyzer were used to issue feedback to the learner. Furthermore, mechanisms of correction by learners were proposed which allows the learner to correct the typed sentence independently, and allows learners to realize what the error is. Arabic ICALL follows the curriculum of Arabic grammar at the Egyptian primary schools. More attention needs to be given to second language learners and further school tests need be applied on real students.

There two priorities for applications involving Arabic language education. Developing a NLP tools that help enable students learn from their mistakes by generating appropriate feedback. Students can rely on software that will be available with a reasonable price instead of spending money in private tutoring. The other one is to teach Arabic for foreigner or second language learners who are interested in either learning Islam or Arabic heritage. From a pedagogical point of view, the proposed rich analyses enable feedback elaboration that helps learners to understand better their knowledge gab.

1. **Dependency Between Technologies**

The following Figure shows the overall architecture of the proposed Arabic ICALL system. 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, syntax analyzer (parser), grammar rules, and lexicon. The feedback component includes an error analyzer that is used to parse ill-formed learner input and to issue feedback to the learner.

In (I)CALL, there are three possible approaches for reacting to the learner's response in order to give appropriate feedback to the learner: *pattern matching-based approach, statistical-based approach,* and *rule-based approach.*

The *pattern matching-based approach* requires that exercise authors enter many different correct and wrong answers with their associated feedback. This is a very tedious and time consuming task yet only provides appropriate feedback when the learner types in one of the expected answers. Thus, this approach requires a great deal of up front teacher knowledge, experience, and effort.

Both *statistical-based approach* and *rule-based approach* give some freedom to the language learners in the way they phrase their answers, while enabling the exercise author to enter only one possible correct answer, thus saving much time compared to the previous pattern matching answer coding approach.

The *statistical-based approach* uses statistical methods to acquire knowledge. Parameters are automatically learned (estimated) from a corpus that is labeled with the properties needed. Both a statistical model and language parameters should be specified by humans. The characteristics of the statistical-based approach are: 1) No strict sense of well-formedness in mind and 2) Have a large parameter space (e.g., 100,000 words using Tri-gram model theoretically requires 105\*105\*105 parameters). The advantages of this approach are that it does not need complex computational models to be established and computation is straightforward.

The disadvantages of this approach are: 1) Cannot restrict computation using heuristics of the linguistic theory, 2) Requires a large amount of data to train the statistical model. Parsers and error diagnosis tools cannot be trained on raw data. The data must be tagged by humans, which is in most cases costly, time consuming, and sometimes ambiguous. Meaning, hand-tagged corpora is very expensive, 3) Even ungrammatical permutation (i.e. sequence) of words of the ill-formed learner's answer are still probable (i.e. have some probability to occur). It is hard to decide whether a string of words is grammatical or not. Actually, strings that are not words at all still will have some probability in the mass probability of the model, 4) The large parameter space of statistical models is a serious problem when decoding (i.e. searching for the most probable structure to assign to a string of words). Statistical models do not distinguish between different forms of words, for example, play, plays, played, playing are not treated as related words originating from the verb play.

The *rule-based approach* provides detailed analysis of the learner's answer using linguistic (morphological and syntactic) knowledge. The characteristics of the rule-based approach are: 1) Has a strict sense of well-formedness in mind, 2) Imposes linguistic constraints to satisfy well-formedness, 3) Allows the use of heuristics (such as a verb cannot be preceded by a preposition), and 4) Relies on hand-constructed rules rather than automatic training from data. The advantages of this approach are that it is easy to incorporate the linguistic knowledge, and it is easy to augment the grammar rules with heuristic rules, which are capable of detecting ill-formed input and providing appropriate feedback. The disadvantage of this approach is that it is not easy to obtain high coverage (completeness) of the linguistic knowledge. However, it could be useful for limited domain where errors in the input can be expected.

It is well-established that feedback is an essential prerequisite for effective learning. Both the pattern matching-based approach and statistical-based approach lack a systematic and automatic way in diagnosing the learner's ill-formed input and providing appropriate feedback. Data collection is costly and time consuming. On the contrary, the rule-based approach has the advantage of providing appropriate feedback because it performs detailed analysis for both well-formed and ill-formed answers. It is conceivable to acquire linguistic knowledge, and to specify linguistic constraints and heuristics. For these reasons, recent works focus on the use of the rule-based approach in developing Arabic ICALL.

**Course Material**

**Learner (Student)**

**Answer/Correction**

**Request**

**Material**

**User Interface**

**(GUI)**

**Material**

**Feedback**

**Request**

**Answer/Correction**

**Feedback**

Overall Architecture of The Proposed Arabic ICALL System

**Sentence Analysis**

**Feedback**

**Teacher (Pedagogical expert)**

1. **Language Resources**
   1. Available Resources (English, Arabic)

There are NLP tools such as morphological analyzers. Also there corpora for Arabic such as LDC. However these are for general NLP tasks.

* 1. Needed Resources (English, Arabic)

There are needs for tools and linguistics resources that can be used for language learning. Current tools are designed for well-formed input while there is a significant needs for tools that analyze ill-formed input in order to give a appropriate feedback to the learner. Also there is an urgent need for a complete computationally erroneous Arabic corpus, the collected for real teaching environment, which can be used to study the educational models.

1. **Strengths, weaknesses, opportunities and threats** 
   1. Strengths

The development of a NLP-based tools and resources for educational purposes will take some efforts to acquire linguistic resources and build components such as question banks, feedback generator, and exam generators. However, the availability of such resources would save lot of efforts and money.

* 1. Weaknesses

Some time, efforts, and investment are required to build a national wide educational system as the NLP tools and resources are unavailable and the current ones are not designed to work with learners input.

* 1. Opportunities

There are lots of benefits that can be gained from teaching Arabic. Student can learn their native language using their own computers. Second language learners, including tourists, can learn Arabic.

* 1. Threats

As there is no such NLP-based intelligent tutoring system is available there will be no threats. An advantage is to be leaders and take a serious step towards our mother tongue.

1. **Suggestions for Survey Questionnaire**

1. **List of people/organizations pioneers in each application area to be targeted by the Survey**

1. **Key persons in each application area (on technical/LR levels)**

1. **Suggestions for Language Resources (specific to the application area) if ALTEC would like to start collection immediately.**

* **Question (item) banking.**
* **Tools for analyzing ill-formed (native and second language) learner input**
* **corpora for erroneous learner answers**
* **marked assessment**

1. **Summary**

NLP tools and resources can be useful for Arabic ICALL for reacting to the learner's response in order to give appropriate 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 their knowledge gab. The market is promising for students looking for an intelligent means for learning Arabic language. The proposed system should be intelligent enough to explain different issues and give the learner feedback on error s/he has made. Another opportunity is for teaching Arabic for people who are looking for learning Arabic to better understand their religion or the Arabic heritage.

REFERENCES

Boytcheva S., Vitanova I., Strupchanska A., Yankova M., & Angelova G. (2004). 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.

Cameron K., editor (1999). CALL-Media, Design, & Applications, Swets & Zeitlinger.

Cushion S. & Hémard D. (2002). Applying New Technological Developments to CALL for Arabic. *Computer Assisted Language Learning (CALL): An International Journal,* Vol. 15, No.5, 501-508, Belgium: SWETS & ZEITLINGER publisher.

Cushion S. & 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, No. 2-3, 259-266, Belgium: SWETS & ZEITLINGER publisher.

Ditters E., Oostdijk N. & Cameron K. (1993). Processing Arabic: Computer Assisted Language Learning. Oxford, England: Intellect Books,.

Gamper J. & Knapp J. (2002). A Review of Intelligent CALL Systems. *Computer Assisted Language Learning (CALL): An International Journal*, Vol. 15, No. 4, 329-342, Belgium: SWETS & ZEITLINGER publisher.

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.

McEnery, T., Baker, J.P., & 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, Belgium: SWETS & ZEITLINGER publisher.

Magdy M. Shalaan K. & Fahmy A. Morphological Analysis of Ill-formed Arabic Verbs in Intelligent Language Tutoring Framework. In FLAIRS-23, The 23rd International FLAIRS Conference, May 19-21, 2010 Daytona Beach, Florida, USA

Maraoui M, Antoniadis G., Zrigui M.: SALA: Call System for Arabic Based. In In IC-AI 2009: Las Vegas, Nevada, USA

Mote N., Johnson L., Sethy A., Silva J. & Narayanan S. (2004). Tactical Language Detection and Modeling of Learner Speech Errors: The case of Arabic tactical language training for American English speakers. In the Proceedings of InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems,Venice. Available at <http://sisley.cgm.unive.it/ICALL2004/link13.htm>

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.

Nielsen H. & 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, Belgium: SWETS & ZEITLINGER publisher.

Othman E., Shaalan K. & 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.

Shaalan K. (2003). Arabic GramCheck: A Grammar Checker for Arabic. Egyptian Informatics *Journal, Faculty of Computers & Information,* , June., Vol. 4, No. 1:94-111, Egypt: *Cairo University.*

Shaalan K., Allam A. & Gomah A. Towards Automatic Spell Checking for Arabic (2003). In Proceedings of the 4th Conference on Language Engineering, Egyptian Society of Language Engineering (ELSE), Oct. 21-22, pp. 240-247, Egypt.

1. CALL is also known as computer-assisted instruction (CAI), computer-aided instruction (CAI), or computer-aided language learning. [↑](#footnote-ref-2)