Developing ARET: An NLP-based Educational Tool Set for Arabic Reading Enhancement

Mohamed Maamouri[1], Wajdi Zaghouani[1], Violetta Cavalli-Sforza[2], Dave Graff[1] and Mike Ciul[1]

[1] Linguistic Data Consortium, University of Pennsylvania, 3600 Market St., Suite 810, Philadelphia, PA 19104.

[2] School of Science and Engineering, Al Akhawayn University, Ifrane 53000, Morocco.

maamouri@ldc.upenn.edu, wajdiz@ldc.upenn.edu, v.cavallisforza@aui.ma, graff@ldc.upenn.edu, mciul@ldc.upenn.edu

Abstract

This paper describes a novel Arabic Reading Enhancement Tool (ARET) for classroom use, which has been built using corpus-based Natural Language Processing in combination with expert linguistic annotation. The NLP techniques include a widely used morphological analyzer for Modern Standard Arabic to provide word-level grammatical details, and a relational database index of corpus texts to provide word concordances. ARET also makes use of a commercial Arabic text-to-speech (TTS) system to add a speech layer (with male and female voices) to the Al-Kitaab language textbook resources. The system generates test questions and distractors, offering teachers and students an interesting computer-aided language learning tool. We describe the background and the motivation behind the building of ARET, presenting the various components and the method used to build the tools.

1 Introduction

Reading is an essential skill for learners of Modern Standard Arabic (MSA). For most of learners it is the most important skill to master in order to ensure success in learning. With strengthened reading skills, learners of Arabic tend to make greater progress in other areas of language learning. Reading should be an active, fluent process that involves the reader and the reading material in building meaning. Often, however, it is not. The average learner's second language reading ability is usually well below that of the first language. This can impede academic progress in the second language. Arabic language teachers and learners face many challenges in the classroom. Teaching students how to utilize the skills and knowledge they bring from their first language, develop vocabulary skills, improve reading comprehension and rate, and monitor their own improvement are just some of the issues that teachers must consider in preparing for an Arabic language reading class. With these issues in mind, we set out to create a web-based service that would provide efficient and pedagogically relevant access to instructional texts in Modern Standard Arabic, with the goal of creating a resource that would serve both instructors and students, by presenting novel modes of information access. We received valuable support from Georgetown University Press, which gave permission for us to use the reading passages from the 3-volume textbook publication *Al-Kitaab* (Al-Batal et al., 2001;2004 and 2006), which is the most popular publication in the USA for teaching Arabic.

2 Motivation

Using technology in classrooms can make the lessons more efficient. There are many technology

The 7th Workshop on the Innovative Use of NLP for Building Educational Applications, pages 127–135, Montréal, Canada, June 3-8, 2012.

tools that can be used in English as a Second Language (ESL) classes to improve foreign students' English and technology skills. According to Wang (2005) there are many advantages integrating technology in classrooms especially for ESL students. To be able to improve their language skills, like writing, reading, listening and speaking, English language learners use pedagogical computer applications to check their work and improve their language skills; they also use web browsers and e-mail to search for information, join in online discussions, publish their work, read technology texts, communicate each other even worldwide. He also says that, "Technology integration in foreign language teaching demonstrates the shift in educational paradigms from a behavioral to a constructivist learning approach" (p. 2). Gone are the days in which learning foreign language vocabulary and grammar rules relied largely on repetitive drills; more and more, foreign language learners are asked to engage directly with authentic materials and take more initiative in their learning. However, finding appropriate, authentic reading materials is a challenge for language instructors. The Web is a vast resource of texts, but most pages are not suitable for reading practice, and commercial search engines are not well suited to finding texts that satisfy pedagogical constraints such as reading level, length, text quality, and presence of target vocabulary. We present a system that uses various language technologies to facilitate the selection , presentation and study of authentic reading materials from the widely used textbook series *Al-Kitaab* (Al-Batal et al., 2001;2004 and 2006). In the next section we review some of the related work. In section 4 we discuss some of the specific challenges faced when learning the Arabic language.

3 Related work

Many studies have shown that an on-line learning environment that supplements classroom instruction with additional study materials at an appropriate level for the learner may enhance language learning and development (Ware, 2004; Chiu et al., 2007; Yuan, 2003; Wang, 2005;). As a result, a number of recent projects have aimed to dynamically provide a supply of accessible authentic texts to language learners by drawing from online resources. WERTi (Meurers et al. 2010) is an intelligent automatic workbook that uses texts from the Web to increase knowledge of English grammatical forms and functions. READ-X (Miltsakaki and Troutt, 2007) is a tool for finding texts at specified reading levels. SourceFinder (Sheehan et al.,2007) is an authoring tool for finding suitable texts for standardized test items on verbal reasoning and reading comprehension. Project REAP (Reader-Specific Lexical Practice) (Brown and Eskenazi, 2004; Heilman et al., 2006) takes a different approach. Rather than teachers choosing texts, in REAP the system selects individualized practice readings from a digital library according to specific lexical constraints. Readings are chosen to contain vocabulary words that a given student needs to learn, while limiting the number of words the student does not know. The choice of texts is therefore driven by a curriculum model, informed by a student model, and constrained by the availability of suitable texts, as described by their text model.

While a user-adapted tool has the potential to better match individual needs, since each student can work with different texts, a drawback of this approach is that instructors may have difficulty coordinating group discussion about readings and integrating the tool into their curriculum. An advantage of a tool containing a search system, however, is that teachers can find texts that match the needs and interests of the class as a whole. While some degree of individualization is lost, the advantages of better coordinated support from teachers and classroom integration are gained. In the early stages of this project, we had planned to use REAP software after adapting it to handle the complex morphology of MSA. Unfortunately, while the system was already being tested in the field, REAP project leaders did not consider the code base mature enough to be released to other research groups. As a result, we chose to develop our own database and access method to texts, foregoing adaptation to individual users.

4 Challenges of Arabic reading

It has never been an easy transition from 'learning to read' to 'reading to learn' for Arabs and other Arabic learners. In Meynet (1971) and according to father Anastase Al-Karmali, a member of the Arabic Language Academy in Cairo, Egypt. "The Arabs study the rules of the Arabic language in order to learn to read, whereas others read in order

to learn …". Indeed, reading in Arabic as a first or second language presents special challenges due to its script and its rich and complex morphology. Also, Arabic texts lack short vowels and other diacritics that distinguish words and grammatical functions. These linguistic complexities result in significant reading difficulties. Typically, Arabic as a second language learners face difficulties in word recognition, word disambiguation and the acquisition of decoding skills, including recognizing letter and word boundaries, decoding unvocalized words and identifying these words. In order to understand Arabic text, the novice reader must learn to insert short vowels and other diacritics based on grammatical rules not yet learned. The ambiguity associated with a lack of diacritization is shown for instance in the lemma علم /Elm/ which has the following nine possible reading interpretations shown in Table 1.

عِلم	'Science, learning'
عَلَم	'flag'
عَلِمَ	3rd P. Masc. Sing. Perf. V. (MSA V. I) 'he learned/knew'
عُلِمَ	3rd P. Sing. Pass. V. (MSA V. I) 'it/he was learned'
عَلَّمَ	Intensifying, Caus. V. (MSA V. II) 'he taught
عَلَّمَ	Causative V. Pass (MSA V. II) 'he was taught'
عِلمُ/عِلمٌ	(NOM Noun + Definite and Indefinite)
عِلمَ	(ACCU Noun + Definite)
عِلمِ/عِلمٍ	(GEN Noun + Definite and Indefinite)

Table 1. Various interpretations for the lemma علم

5 The Arabic reading enhancement tools

To address these challenges, we developed an Arabic Reading Enhancement Tool (ARET) for classroom use with support from the U.S. Department of Education's International Research Study Program (IRS). The ARET tool is rather similar in intent to the foreign language learning tool, GLOSSER-RuG built by Nerbonne and Smit (1996) for Dutch, but targets explicitly the particularities of MSA. ARET has two subparts tools : the Arabic Reading Facilitation Tool (ARFT) and the Arabic Reading Assessment Tool (ARAT). A major achievement of this project was to create a collection of fully annotated texts for learners of Arabic, using materials included in an authoritative textbook series that spans several competence levels. In this section, we describe the creation, structure and content of the Arabic corpus/lexicon database, and then describe the ARFT and ARAT tools in more detail.

5.1 The Al-Kitaab corpus database

The ARET system uses the full text of Arabic reading passages from the Georgetown University Press *Al-Kitaab* textbook series, which represents a 60,000 word corpus. Each passage was submitted to a combined automatic/manual annotation process in order to create a version of the text that was completely diacritized and thoroughly segmented and labeled to identify all morphemes for each word, including their part-of-speech labels and English glosses.

We first applied the Standard Arabic Morphological Analyzer (SAMA) (Maamouri et al., 2010), to enumerate all possible solutions for each word token in a given passage. The entire passage, with the full set of possible SAMA solutions for each word token, was then presented to a native Arabic speaker experienced in the morphological analysis of MSA, and their task was to select the particular SAMA solution for each word based on their understanding of the context; where necessary, the annotator would manually edit the details of POS tags or glosses to fill gaps in SAMA's coverage of the vocabulary. This is a standard approach used in the annotation of numerous Arabic text corpora, including the Arabic Treebank Project (Maamouri and Bies 2004). As described in section 5.2, the resulting annotation was fully reviewed by expert Arabic linguists using our reading facilitation tool, to identify and repair errors.

A relational database was created to store the corpus and annotations. Separate tables were used to enumerate (a) the reading passages (keeping track of the book volume, chapter and page number of each passage), (b) the sequence of sentences in each passage, (c) the word token sequence for each sentence, (d) the inventory of distinct word types (i.e. orthographic word forms with their context-dependant analyses), and (e) the inventory of distinct "headwords" (lemmas) and affix morphemes (clitics).

Using this relational table structure, a full passage could be assembled for display by querying

for the sequence of sentences and the word tokens for each sentence. The information returned by the query could include, for each word token, the original and/or diacritized spelling, and an index for looking up the context-dependent morphological analysis plus gloss for the token. This in turn also provided access to a dictionary entry for the lemma from which the token was derived. Table 2 summarizes the contents of the database. The number of distinct lemmas refers to the number of citation forms for content words (nouns, verbs, etc) that are referenced by the all the inflected stems found in the reading texts; the number of glossary entries refers to the manually edited dictionary descriptions for lemmas / citation forms, including their consonantal roots. In cases where a lemma does not have a corresponding glossary entry, the fully-detailed morphological analysis provides an English gloss (but not the root) for each word token containing the lemma.

Type	No. of Entries
Sentences, titles and sub-headings	3,692
Arabic word tokens	53,411
Distinct undiacritized Arabic orthographic forms	17,209
Distinct diacritized orthographic forms	20,725
Distinct morphology/POS/gloss annotations on word forms	22,304
Distinct clitic and inflected-stem morphemes	16,774
Distinct lemmas	6,829
Glossary entries for lemmas	3,436

Table 2. Corpus quantities in ARET database

5.2 The Arabic reading facilitation tool

The Arabic Reading Facilitation Tool (ARFT) provides the user with direct access to the *Al-Kitaab* text corpus, organized by volume, chapter and page number. In addition to presenting the full text for a given passage, the user can click on any word in the passage to bring up in a side-bar the full morphological analysis and gloss for the word in that context, along with a glossary entry for the associated lemma, and a summary of other Arabic citation forms that are related by root. Two other important functions are also provided: (a) toggling the presence vs. absence of all diacritic marks in the full display of the reading passage, and (b) the ability to view a concordance of all occurrences for any selected word. The tool also provides a "tool-tip" pop-up window whenever the mouse cursor hovers over an Arabic word in the text passage; if the page is showing undiacritized text, the pop-up shows the diacritized form of the word, and vice-versa. This is a very useful feature for the new learners of the Arabic language.

As soon as the annotated version of the corpus was loaded into the database, there was a sustained effort involving native Arabic speakers and Arabic faculty to carefully review the database content, as displayed by the ARFT, and validate it against the original textbook content. This effort involved numerous repairs of all sorts that stemmed from all stages of corpus preparation: typing mistakes from the original keyboarding of the text, problems in morphological annotation, and difficulties in the loading of the tables. Customized tools and procedures were developed to facilitate the updates that were needed to apply all the corrections directly to the database.

A glossary for use in the ARFT was added to the database, with the relational linkage needed to support glossary lookups triggered by the user clicking on any word in a text passage. The word-to-glossary relation is based on the "lemma_ID" of the stem in each word. The lemma_ID is a string identifier assigned by the Standard Arabic Morphological Analyzer (SAMA), which was used for the morphological annotation of the entire corpus; all verbs in a given conjugation paradigm share the same lemma_ID, as do all nouns or adjectives in a given declensional (case) paradigm, so every distinct inflected form of a noun, adjective or verb is linked by the annotation to its corresponding glossary entry. The glossary table (with indexing by Semitic root) was a special, additional annotation specifically for ARFT, so not all lemmas were covered in the glossary; when a term not in the glossary is clicked, the side-bar display area in the ARFT shows the message "Refer to Morphology Information"; the morphology information is the full set of annotation data for each word based on SAMA, and this always includes an English gloss for the stem (except in the case of proper nouns, which always have "Proper Noun" as their part-of-speech label).

The ARFT is intended for use with a modern web browser over a reasonably fast internet con-

nection. The tool has a flexible and intuitive web interface to navigate the texts via several key features:

1. Source Panel, featuring *Al-Kitaab* text
2. Highlighted Sentence
3. Highlighted Word
4. Audio Player for highlighted sentence
5. Audio Player for highlighted word
6. Morphological Data Panel
7. Lexical Data Panel
8. Tabbed browsing for convenient access to multiple screens

Figure 1. below illustrates an example of the tool using a passage of text from Al-Kitab Volume 2, Page 61.

Figure 1. Arabic Reading Facilitation Tool featuring function labels

5.3 The Text to speech module

An Arabic Text-to-Speech technology module was licensed from RDI[1]. This technology has been used to add an audio feature to the ARFT, and can be used to render audio of arbitrary Arabic text. So the users will be able to listen to individual words or passages of text spoken by a high quality synthesized voice. The RDI module, reads text files or literal text in Windows Arabic encoding and generates WAV audio data either as files or for direct output to an audio device. It has a C++ API that may be employed in Microsoft Visual Studio. The voice rendering quality is excellent. Moreover, the module analyzes diacritized or undiacritized Arabic text to determine pronunciation, rhythm and inflection of speech. Many variables of speech production can be controlled, most significantly the gender of the speaker. We developed a simple console-based executable that reads a list of Arabic text files and generates a WAV file of speech corresponding to each one, using a male voice, female voice, or one of each.

5.4 The Arabic Reading Assessment Tool (ARAT)

In order to support the creation of tests and quizzes for specific Arabic reading skills the Arabic Reading Assessment Tool (ARAT) has been built around an existing open-source web application framework called Moodle (http://moodle.org). This framework was developed as a "Content Management System", and provides built-in support for many of the 'infrastructure' functions that ARAT would need, including: registration of faculty and student user accounts; creation of courses with schedule plans and content-based resources; creation, presentation and scoring of tests and quizzes; and overall record-keeping of resources, activities and test scores. Custom software modules were developed to augment the Moodle code base in order to provide functions that are specific to the ARAT:

- communicating with and importing data from the annotated *Al-Kitaab* passage database;
- defining specialized question types (the first three types described below) based on annotations in the database, such that answers to the questions can be scored automatically by reference to the corpus annotations.

The three types of annotation-based questions were defined and implemented in the prototype ARAT:

- Cloze-Test Question: given a reading passage in Arabic, one or more words are chosen as test items and are replaced in the text by an underlined empty slot; the student is given a multiple-choice question to identify the correct Arabic word to fill each slot.
- English Gloss Question: given a reading passage, one or more words are chosen as test items and highlighted in the text; the student is given a mul-

[1]<http://www.rdi-eg.com/Technologies/speech.htm>

tiple-choice question to identify the correct English gloss for each test word.

- Case-Ending Question: given reading passage, one or more nouns and/or adjectives are chosen as test items and highlighted in the text; the student is given a multiple-choice question of the six possible cases in Arabic to identify the correct case ending for each test word. Mood ending could also be considered for verbs.
- Yes/No questions: these are fully developed by teachers, who must enter questions and answers into the program in order to have Moodle give the student/teacher the appropriate final scores and correct answers feedback.

The implementation allows an instructor to select what text passage to use for a given quiz, and also allows for either manual and automatic selection of words to use as test items from the text, as well as either manual or automatic selection of distractor items for the Cloze and Gloss tests. By providing automatic selection of test items and distractors based on available annotations in the corpus database, ARAT allows a student to practice each task any number of times on a given text passage, be challenged by novel questions on each attempt, and receive a tally of right and wrong answers, without the instructor having to create or score each attempt as shown in Figure 2.

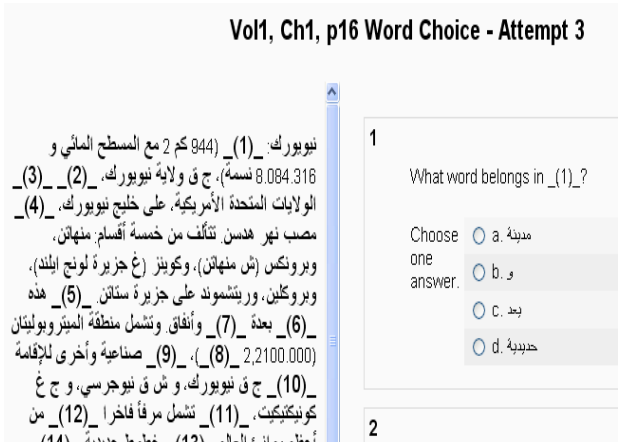

Figure 2. A sample question created with ARAT

5.5 The test set creation procedure

The procedure for creating a test set within ARAT breaks down to the following 'top-level' steps:

1. Provide or select a text passage to be used as the source from which test questions are derived.

2. For questions that will be based on specific word tokens in the text, identify the tokens that will be basis for test questions; these token-specific questions will always involve a particular task with a multiple-choice response, so for each selected token: select the task (word choice, gloss choice, case-ending), identify a correct answer and provide or select a set of three distractors.

3. For questions not based on specific tokens, the instructor must supply the following: prompting text for the question, the type of response (y/n, t/f, type-in, multiple choice) and the correct answer (and three distractors for multiple choice). Figure 3 shows the test set main screen.

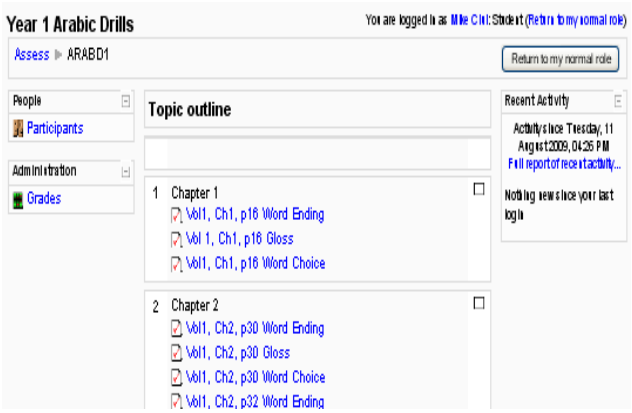

Figure 3. Test set main screen.

6 Classroom usage and tool evaluation

The ARFT was presented to Arabic faculty at the University of Pennsylvania; the tool was announced in Arabic courses and students were asked to use it. Several lists of student enrollments for many Arabic courses have been imported into the Moodle-based system.

An informal evaluation was also performed, in the Summer of 2010, with Arabic instructors teaching in the ARAbic and North African Studies (ARANAS) program at Al Akhawayn University, in Ifrane, Morocco. Unfortunately, due to the very rushed schedule and time pressure that instructors work under during this intensive program, the tools did not receive the desired attention. Only a handful of instructors actually explored the tools. Two

instructors filled out an evaluation questionnaire concerning various aspects of the tools and their use of computer technology for language teaching in general. The feedback was generally positive and included some detailed suggestion for improving the tools; they also revealed some issues with inconsistent response time (partly due to the network infrastructure of the university at that time) and ease of use (for non technology-savy instructors). The biggest obstacles to using the tools, however, appeared to be lack of time on the part of the instructors to acquire sufficient familiarity with the tools and devise effective ways of introducing them in the curriculum. We are investigating the possibility of using the tools with exchange students during the regular academic year, even though the numbers in Arabic classes at all levels is much lower than in the Summer program.

Recently, the use of the ARFT and its companion the ARAT has been mandated by the Arabic Section at the University of Pennsylvania and we hope that a more consistent use is going to be made. As of now, 118 students are registered representing four 1st Year classes (total: 63 students), two 2nd Year classes (total:3 students), one 3rd Year class (total: 13 students) and One 4th Year class (11students).At this point, the tool impact on the classroom has not been evaluated, but it is in our future plans to do a comprehensive classroom evaluation of the tool.

As part of the effort to introduce the ARFT and the ARAT to faculty, we obtained three short reading passage texts, totaling 1022 Arabic word tokens, selected by a faculty member from news sources. These were submitted to annotation to disambiguate and diacritize the content based on SAMA analysis, just as was done for the *Al-Kitaab* passages. The annotated texts have been added into the database corpus and are available for use in the ARAT, but are not accessible for general browsing via the ARFT. The annotation and database import went quickly, demonstrating that these procedures have matured, and providing resources for building quizzes and tests based on materials that are 'unseen' by students who use both the ARFT and the ARAT.

7 Conclusion

We have described computational tools and linguistic resources that enable students to enhance their Arabic reading skills by helping them with the difficulties they face in word recognition, word disambiguation and general decoding skills during the Arabic reading process. These computational tools and resources provide the needed correct and meaningful vocalizations by using natural language processing (NLP) technologies namely a Standard Arabic Morphological Analyzer (SAMA), a concordance, a Text-to-Speech module and various interfaces. The time gained by students who use our Reading Enhancement Tools could be put to good use in the current ASL (Arabic as a Second Language) classroom which, following the ACTFL proficiency movement puts a primary emphasis on communication with less concern for accuracy as reflected in morphology or syntax, particularly at the initial stages of ASL learning. We reiterate at this point that our choice of the GUP *Al-Kitaab* textbook series was not fortuitous. We could have chosen any other pedagogical text but Al-Kitaab distinguishes itself by being widely used in the United States and abroad, and providing an extensive curriculum with a wide variety of texts. We are thankful that GUP gave us permission to use this resource, as it enabled us to create a tool that can accompany many English-speaking students studying MSA in many classrooms around the world.

In addition to answering learners' reading needs in MSA, our efforts went beyond the specificities of this language by allowing us to demonstrate that our tools and the methodology we followed was in fact 'portable' to other languages which had a morphologically complex nature such as, for instance, the Nahuatl Learning Environment (NLE) project based on the ARET infrastructure [2]. Future efforts will continue experimentation of the use of available and robust Arabic NLP technologies to extend the enhancement of Arabic reading to better understanding of authentic reading text that the reader could download from the Internet for instance. Progress in that direction is desirable and possible because it would increase the motivation of Modern Standard Arabic learners and will boost access by students and other professionals to authentic real world language text in new genres and topics. In this way, the contribution of NLP tech-

[2] The Nahuatl learning tool project prepared by Jonathan Amith (n.d) and a team of Nahuatl speakers can be accessed online through a Beta version of the Nahuatl Learning Environment at the LDC : http://nahuatl.ldc.upenn.edu/.

nologies to the teaching and learning of languages may become more significant and more compelling to all concerned, teachers, learners and computer NLP specialists alike.

Acknowledgements

We gratefully acknowledge the sponsorship of the U.S. Department of Education, whose International Research Study (IRS) Grant No. P017A050040-07-05 supported our work on this project. The views, opinions and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the U.S. Department of Education's International Research Study program. We also acknowledge the help and support of Georgetown University Press who allowed us to use their Al-Kitaab series as a testing ground for our tools. Thanks and appreciation go Professor Roger Allen and his team of Arabic teachers at the University of Pennsylvania for their warm reception of our tools in their teaching structure. Thanks go finally, to all the programmers and annotators who worked on the project. They are numerous and we cannot give them all the credit they deserve but without them our achievement would not have been so significant.

References

Mahmoud Al-Batal, Kristen Brustad & Abbas Al-Tonsi. 2006. *Al-Kitaab fii tacallum al-cArabiyya*, Volume II (with DVDs, Second Edition).Washington, D.C.: Georgetown University Press, 2006.

Mahmoud Al-Batal, Kristen Brustad & Abbas Al-Tonsi. 2004. *Al-Kitaab fii tacallum al-cArabiyya*, A Textbook for Beginning Arabic, Volume I (with DVDs, Second Edition). Washington, D.C.: Georgetown University Press, 2004.

Mahmoud Al-Batal, Kristen Brustad & Abbas Al-Tonsi. 2001. *Al-Kitaab fii tacallum al-cArabiyya*,Volume III.Washington, D.C.: Georgetown University Press, 2001.

Jonathan Amith. n.d. *Nahuatl Learning Environment.* Available online at : http://nahuatl.ldc.upenn.edu/.

Jon Brown and Maxine Eskenazi. 2004. Retrieval of authentic documents for reader-specific lexical practice. In *Proceedings of* InSTIL/ICALL Symposium 2004. Venice, Italy.

Tsuo-Lin Chiu, Hsien-Chin Liou and Yuli Yeha. 2007. A study of web-based oral activities enhanced byautomatic speech recognition for EFL college learning. Computer Assisted Language Learning, 20 (3), 209–233.

Michael Heilman, Kevyn Collins-Thompson, Jamie Callan, and Maxine Eskenazi. 2006. Classroom success of an Intelligent Tutoring System for lexical practice and reading comprehension. In *Proceedings of* the Ninth International Conference on Spoken LanguageProcessing. Pittsburgh, PA.

Mohamed Maamouri, David Graff, Basma Bouziri, Sondos Krouna, Ann Bies and Seth Kulick. 2010. Standard Arabic Morphological Analyzer (SAMA) Version 3.1. Linguistic Data Consortium, Catalog No.: LDC2010L01.

Mohamed Maamouri and Ann Bies. 2004. Developing an Arabic Treebank: Methods, Guidelines, Procedures, and Tools. In *Proceedings of* the Workshop Computational Approaches to Arabic Script-based Languages. Pages 2-9./20th International Conference on Computational Linguistics/. COLING Geneva, Switzerland.

Detmar Meurers, Ramon Ziai, Luiz Amaral, Adriane Boyd, Aleksandar Dimitrov, Vanessa Metcalf, Niels Ott. 2010. Enhancing Authentic Web Pages for Language Learners. In *Proceedings of* the 5th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT 2010, Los Angeles.

Roland Meynet. 1971. *L'écriture arabe en question: les projets de l'Académie de Langue Arabe du Caire de 1938 à 1968*.Beirut: Dar el-Machreq, 1971. 142 pp

Eleni Miltsakaki and Audrey Troutt. 2007. Read-X: Automatic Evaluation of Reading Difficulty of Web Text. In *Proceedings of* E-Learn 2007, sponsored by the Association for theAdvancement of Computing in Education. Quebec, Canada.

John Nerbonne and Petra Smit. 1996. GLOSSER-RuG: in Support of Reading. In *Proceedings of* the 16th International Conference on Computational Linguistics (COLING 1996).

Kathleen M. Sheehan, Irene Kostin, Yoko Futagi. 2007. SourceFinder: A Construct-Driven Approach for Locating Appropriately Targeted Reading Comprehension Source Texts. In *Proceedings of* the SLaTE Workshop on Speech and Language Technology in Education.Carnegie Mellon University and International Speech Communication Association (ISCA).

Li Wang. 2005. The advantages of using technology in second language education. T.H.E. Journal, 32 (10), 1-6.

Paige D. Ware. 2004. Confidence and competition online: ESL student perspectives on web based discussions in the classroom. Computers and Composition, 21, 451–468.

Yi Yuan. 2003. The use of chat rooms in an ESL setting. Computers and Composition, 20, 194–206.