TOOLS

Arabic Reading Enhancement Tools

Challenges of Arabic Reading

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 in the lemma علم which has nine possible reading interpretations:

- عِلْم - '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'
- عُلِّمَ - Causative V. Pass (MSA V. II) 'he was taught'
- عِلْمُ/عِلْمٌ - (NOM Noun + Definite and Indefinite)
- عِلْمَ - (ACCU Noun/Definite)
- عِلْمِ/عِلْمٍ - (GEN Noun + Definite and Indefinite)

The Arabic Reading Process

To address these challenges, LDC has developed the Arabic Reading Enhancement Tools (ARET) with support from the U.S. Department of Education's *International Research Study Program* (IRS), Grant No. P017A050040-07-05. These tools include:

- Arabic Reading Facilitation Tool (ARFT)
- Dictionary lookup
- Morphological Analysis
- Concordance output showing all occurrences of a given word (with links to full texts)
- Arabic Reading Assessment Tool (ARAT)

Arabic Reading Facilitation Tool (ARFT)

The Arabic Reading Facilitation Tool gives the user multiple views of a fully annotated Arabic text and provides the ability to:

- Show and hide diacritic marks
- Listen to male or female vocalizations of the text using an Arabic Text-to-Speech synthesizer
- View lexical and/or morphological data about any highlighted word
- Search a concordance for all occurrences of a word in the reading text

Arabic Reading Facilitation Tool featuring function labels

Tool Features

The ARFT has an easy-to-use web interface that uses flexible, intuitive programming technologies, specifically AJAX, to allow the user to navigate the texts using 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

Linguistic Data Consortium 3600 Market St., Suite 810, University of Pennsylvania, Philadelphia, PA 19104-2653 USA
Telephone: +1.215.898.0464 • Fax: +1.215.573.2175 • ldc@ldc.upenn.edu • www.ldc.upenn.edu

Annotated Text

LDC ARET use the complete texts of the Georgetown University Press *Al-Kitaab* Arabic textbook series, which represents a 60,000 word corpus. These texts were automatically analyzed and annotated to provide full diacritization and morphological data. The annotation was put through a human quality control pass and validation for a 'Zero Error' diacritized text.

Why *Al-Kitaab*?

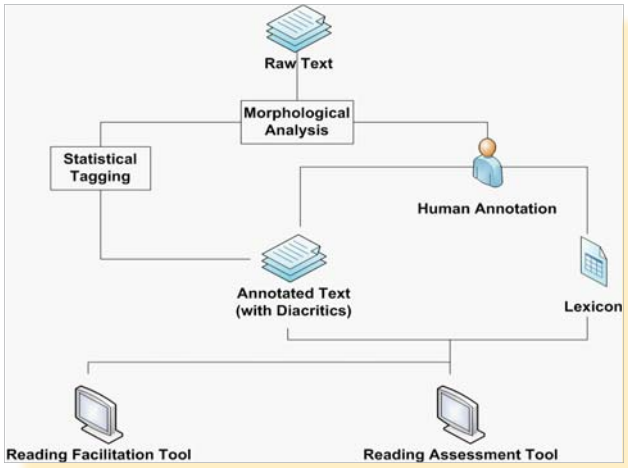
Al-Kitaab is the most commonly used Arabic textbook in both the US and abroad. These texts enabled LDC to provide a 'proof of concept' to show that the Arabic reading tools can be ported to other languages. LDC believes that its linguistic tools should not be linked to a given teaching methodology, but rather to learners' and educators' needs.

Georgetown University Press permitted LDC's use of these publications.

Tool Creation

The tool development process leveraged LDC's expertise in NLP (natural language processing) tool creation and harnessed the powerful contribution of cutting edge NLP technology in order to provide most of the required annotation. LDC also utilized their extensive and expanding collection of Arabic language resources, especially the LDC Standard Arabic Morphological Analyzer (SAMA) and the Penn Arabic Treebank.

LDC ARET can be used by all age groups, from K-12 through higher education.

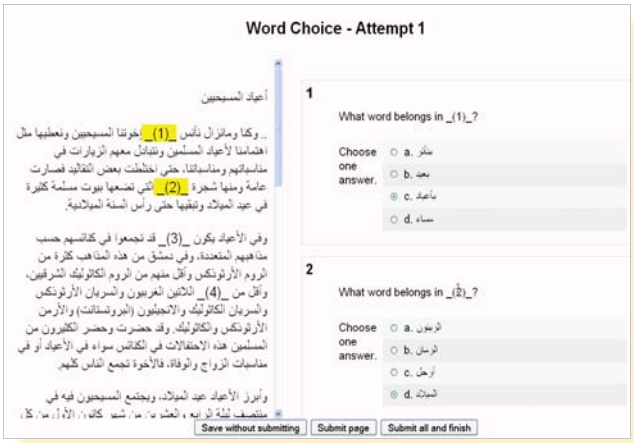


Flow chart depicting tool development

Arabic Reading Assessment Tools (ARAT)

The Arabic Reading Assessment Tools allow users to practice with both randomly- and manually- (i.e., instructor) generated tests on the provided texts. Four types of tests are available:

- English Gloss questions - multiple choice, identify the correct English word
- Cloze Test questions - multiple choice, identify the correct Arabic word
- Case/Mood Ending questions - identify the correct case/mood ending for each targeted word
- True-False questions (manually-generated only)



Screenshot of a Reading Assessment test

LDC Tools and the Foreign Language Classroom

Because of the focus on oral communication, reading comprehension may receive little classroom attention. As a result, students may suffer from a lack of reading comprehension.

LDC ARET shift the focus from textbook and teacher-centered teaching methodologies to home-based independent learning enrichment. Improving reading skills saves time for classroom-based communicative activities and linguistic exercises. Reading practice in turn strengthens the acquisition of lexical skills.

Tool Website and Documentation

LDC Arabic Reading Enhancement Tools and all appropriate documentation, including authorship information, can be accessed at:
http://web1.ldc.upenn.edu/Projects/art/
Contact: Dr. Mohamed Maamouri (Project PI)
maamouri@ldc.upenn.edu