



South Asia Language Resource Center

SALRC GRANT PROPOSAL APPLICATION FORM 2005-2006

Proposal title
Digital Course Pack: Urdu Ghazal

Language(s)
Urdu

Project period (Earliest start date is 8/15/2005;
project must be completed by 8/14/2006)
8/15/2005-8/14/2006

Is this proposal an extension or expansion of a
previously SALRC funded project?
No

Faculty PI
Frances Pritchett

Department
MEALAC

Office address
602 Kent Hall, Mail Code 3928
1140 Amsterdam Avenue
City State Zip code
New York, NY 10027

Have you previously received SALRC funding
for an unrelated project?
No

Requested funding amount
\$ 16,800

Institution
Columbia University

Office telephone FAX
212-854-4370 212-854-5517

E-Mail
fp7@columbia.edu



South Asia Language Resource Center

Faculty co-PI (if applicable)
A. Sean Pue

Department
MEALAC

Office address
602 Kent Hall, Mail Code 3928
1140 Amsterdam Avenue
New York, NY 10027

Institution
Columbia University

Office telephone: 646-220-4335 FAX 212-854-5517

E-Mail
asp49@columbia.edu

Faculty co-PI (if applicable)

Department

Office address

City State: Zip Code

Institution

Office telephone: FAX

E-Mail

Description of proposal (Please do not exceed three pages.):

You must respond to the points covered in the RFP including technology, timeline, budget, pedagogical approach and the unique aspect of this project compared to existing materials.



South Asia Language Resource Center

A. Project Overview

This project will consist of an on-line Urdu course reader which can be used alone as the reader for a semester-long advanced Urdu class or as supplementary material for a first or second-year class. The interactive on-line reader will aid students in orthography, pronunciation, vocabulary, grammar, and cultural and literary context. A printable version will also be available in order to accommodate classroom work.

B. Reader Contents

The reader will consist of eighteen ghazals by ten poets (Vali, Dard, Mir, Jur'at, Atash, Momin, Zauq, Zafar, Ghalib, and Iqbal). These poems have been selected for pedagogical appropriateness. They will be arranged chronologically and, as far as possible, in order of difficulty.

C. Reader Assistants

The on-line versions of the Urdu text will be interactive. Framing the text of the poems will be menus offering the student the option of several "Assistants":

1. *Reading Assistants*, which can be turned on and off, will aid the student in orthography and pronunciation when the student moves the mouse over any word in the text. These will include showing the letters separately, showing the words in transliteration, and showing the words in IPA.
2. *Vocabulary Assistants* will also be available for certain words, which will be highlighted when the student moves the mouse over them. Clicking these words will display the vocabulary information for the word or word phrase, including case, origin, and definition.
3. *Grammar Assistants* will follow certain lines of the text. These which will review elements of Urdu grammar present in the line.
4. *Context Assistants* will also follow certain lines of the text and aid in literary and cultural context.
5. There will also be an icon to "Recite," which will play an audio file of a native speaker reading the verse.

D. Pedagogical Merits

There are a number of existing analog texts for developing reading skills in Urdu. The best reader for the ghazal, M.A.R. Barker's *Classical Urdu Poetry* [1977], is out-of-print and unavailable. Unlike Barker's reader, we will not provide a translation of the texts, as we find that these offer the students too much of a crutch. In place of translations, this digital environment will provide the students with alternative tools to read through the texts by themselves before discussing them in more detail in the classroom. The Urdu orthography is difficult, and it takes a long time before students can read fluently and quickly.

While there are no existing digital readers for Urdu, and thus no way to evaluate their merits, we expect that a digital environment will help students develop reading skills efficiently. With students more and more tied to multimedia forms, digital learning environments make particularly good sense for today's undergraduates. But this project also draws on the more old-fashioned concept of individual learning, allowing students to test their abilities on an individual level. This is particularly apt for the study of lesser-commonly taught languages, where classes often accommodate students of widely divergent skill levels and linguistic backgrounds.

The Urdu ghazal itself has a number of pedagogical benefits for all-around competency in Urdu. It is the most culturally prestigious, ubiquitous, memorized and memorizable literary form in Urdu. It is the key to much literary and popular cultural imagery, including the Hindi film and other popular musical genres. Because beginning Urdu students are constrained by limited reading skill, the brevity and memorizability of ghazal couplets are particularly helpful. Despite its reputation as an exquisite literary genre, the ghazals chosen for this reader employ common grammatical constructions and vocabulary useful to all students. Their recitation also provides good oral practice.



South Asia Language Resource Center

This reader has proven to be successful for many years in its analog form. We will test the digital reader in second and third year Urdu classes at Columbia University in order to receive feedback on the merits of the digital environment. Using the SALRC network, we will encourage colleagues to assign these exercises in their classes and provide feedback.

E. Web Deployment and Accessibility

This project will produce a completed course packet which will not require server-side operations and can therefore be hosted at Columbia University, SALRC, and at other sites. This project will conform to World Wide Web Consortium (W3C) standards, and will therefore be accessible for years to come. It will also be available as a packet which can be downloaded by students for use off-line.

The packet will consist entirely of XHTML files and will conform to the Web Accessibility Initiative. The SALRC font recommendations will be the default for the Urdu and diacritics scripts. Users will also have the option of selecting their own fonts. (Settings will be stored in cookies, and Javascript will adjust the fonts on the CSS style sheets as the pages load). Audio files will be deployed as MP3s and recorded in the soundproof room at the Language Resource Center at Columbia University. PDF files with embedded fonts will be provided for the printable version.

F. Technology Overview

The texts in the reader will be encoded into a rules-based markup language capable of rendering the following:

Markup language	-->	Urdu script	آب
	-->	Transliteration	āb
	-->	IPA	ab

We have developed a prototype for the markup language that produces the Urdu script and transliteration. Some adjustments and supplementation will be necessary in order to accurately render IPA.

The texts will be stored within a local MySQL database which will be used to generate the XHTML files. The engine for the project will be coded in Java.

The technology used in this reader can easily be extended to work with other genres of Urdu texts. Furthermore, as the markup language will work from a set of XML rules, the technology and its developing environment can later be reworked for use with other South Asian languages.



South Asia Language Resource Center

G. Project Requirements and Timetable

Task	Hours
An “analog” version of this reader already exists as Notabene files. It must be converted into the markup language and put into a MySQL database.	40
The vocabulary for this reader must be converted from Notabene into the markup language, edited, and put into a MySQL database.	30
Java functions to convert the markup language into Unicode Urdu script, diacritics, and IPA must be coded.	40
The texts must be proofread in their different forms and the markup language finalized.	40
Grammatical and cultural context information must be written. These will then be put into a MySQL database and references to them will be mapped onto the Urdu texts.	80
A Java Swing environment must be coded for editing and manipulating the texts as well as their “assistant” information.	100
Code must be written that works within the editing environment to map vocabulary assistants onto the Urdu texts. This process will be mostly automated, by querying a database, but will also allow for manual mapping and unmapping.	80
All of the texts and vocabulary words will be read by a native speaker, preferably with experience in radio, and recorded at the Columbia University Language Research Center. These will then be linked to the Urdu texts in the database.	30 (PI) 20 (Voice Provider)
Java code must be written to generate the XHTML web pages.	40
Code must be written to generate vocabulary lists.	20
The introductory front material and help files must be written.	10
The printable PDFs of the course pack must be generated.	10
At this point, the web pages will be available for testing by the Urdu community and by language teachers, and announcements will be sent out to relevant listservs.	
Final beta testing, revisions, and packaging must be completed, and the web pages must be deposited. We can post the web pages at the Language Resource Center at Columbia University, and we would be happy if SALRC could also host the course pack.	20
Total hours:	560
Total Budget (at \$30/hr):	\$16,800