

U.S. Department of Education

Washington, D.C. 20202-5335



APPLICATION FOR GRANTS UNDER THE

INTERNATIONAL RESEARCH AND STUDIES - INSTRUCTIONAL MATERIALS

CFDA # 84.017A

PR/Award # P017A060058

Grants.gov Tracking#: GRANT00057168

Closing Date: NOV 16, 2005

****Table of Contents****

Forms

1. Standard Face Sheet (ED 424)	e1
2. Standard Budget Sheet (ED 524)	e2
3. SF 424B - Assurances Non-Construction Programs	e4
4. Disclosure of Lobbying Activities	e6
5. 427 GEPA	e7
Attachment - 1	e9
6. ED 80-0014 Certification	e10
7. ED 80-0013 Certification	e11

Narratives

1. Project Narrative - (Abstract Narrative...)	e12
Attachment - 1	e13
2. Project Narrative - (Project Narrative...)	e14
Attachment - 1	e15
Attachment - 2	e45
3. Project Narrative - (Other Narrative...)	e46
Attachment - 1	e47
4. Budget Narrative - (Budget Narrative...)	e117
Attachment - 1	e118

This application was generated using the PDF functionality. The PDF functionality automatically numbers the pages in this application. Some pages/sections of this application may contain 2 sets of page numbers, one set created by the applicant and the other set created by e-Application's PDF functionality. Page numbers created by the e-Application PDF functionality will be preceded by the letter e (for example, e1, e2, e3, etc.).

Application for Federal Education Assistance



U.S. Department of Education
Form Approved
OMB No. 1890-0017
Exp. 06/30/2008

Applicant Information

1. Name and Address

Legal Name: CRCL Inc.
Address: 820 Calle Pluma

San Clemente
City

CA
State

2. Applicant's D-U-N-S Number

3. Applicant's T-I-N

4. Catalog of Federal Domestic Assistance #:

Title: International Research and Studies - Instructional Materials

5. Project Director: Doug Cooper

Address: 820 Calle Pluma

San Clemente CA 92674 -
City State ZIP Code + 4

Tel. #: (949) 496-5195 Fax #: () -

E-Mail Address: DOUG@TH.NET

Organizational Unit

Orange 92673 - 2719
County ZIP Code + 4

6. Novice Applicant Yes

7. Is the applicant delinquent on any Federal debt? No
(if "Yes," attach an explanation.)

8. Type of Applicant (Enter appropriate letter in the box.)

- A State
- B Local
- C Special District
- D Indian Tribe
- E Individual
- F Independent School District
- G Public College or University
- H Non-Profit College or University
- I Non-Profit Organization
- J Private, Profit-Making Organization
- K Other (Specify):

9. State Application Identifier:

Application Information

10. Type of Submission:

--Application

Non-Construction

11. Is application subject to review by Executive Order 12372 process?

No - Program is not covered by E.O. 12372

13. Are any research activities involving human subjects planned at any time during the proposed project period? No (Go to item 14.)

13a. Are **all** the research activities proposed designated to be exempt from the regulations?

Yes (Provide Exemption(s)#):

No (Provide Assurance #, if available):

12. Proposed Project Dates:

Start Date:	End Date:
<u>10/01/2006</u>	<u>09/30/2009</u>

14. Descriptive Title of Applicant's Project:
SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition in Complex-Script Languages

Estimated Funding

Authorized Representative Information

15a. Federal	\$	178,481	.00
b. Applicant	\$	0	.00
c. State	\$	0	.00
d. Local	\$	0	.00
e. Other	\$	0	.00
f. Program Income	\$	0	.00
g. TOTAL	\$	178,481	.00

16. To the best of my knowledge and belief, all data in this preapplication/ application are true and correct. The document has been duly authorized by the governing body of the applicant and the applicant will comply with the attached assurances if the assistance is awarded.

a. Typed Name of Authorized Representative
Doug Cooper

b. Title
Executive Director

c. Tel. #: (949) 496-5195 Fax #: () -

d. E-Mail Address: DOUG@TH.NET

e. Signature of Authorized Representative



U.S. DEPARTMENT OF EDUCATION
BUDGET INFORMATION
NON-CONSTRUCTION PROGRAMS

OMB Control Number: 1890-0004

Expiration Date: 06/30/2005

Name of Institution/Organization:
 CRCL Inc.

Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION A - BUDGET SUMMARY
U.S. DEPARTMENT OF EDUCATION FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	\$ 77,598	\$ 80,702	\$ 83,930	\$ 0	\$ 0	\$ 242,230
2. Fringe Benefits	\$ 26,383	\$ 27,439	\$ 28,536	\$ 0	\$ 0	\$ 82,358
3. Travel	\$ 5,000	\$ 5,000	\$ 5,000	\$ 0	\$ 0	\$ 15,000
4. Equipment	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
5. Supplies	\$ 3,600	\$ 0	\$ 0	\$ 0	\$ 0	\$ 3,600
6. Contractual	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
7. Construction	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
8. Other	\$ 65,900	\$ 68,900	\$ 65,900	\$ 0	\$ 0	\$ 200,700
9. Total Direct Costs (lines 1-8)	\$ 178,481	\$ 182,041	\$ 183,366	\$ 0	\$ 0	\$ 543,888
10. Indirect Costs*	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
11. Training Stipends	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
12. Total Costs (lines 9-11)	\$ 178,481	\$ 182,041	\$ 183,366	\$ 0	\$ 0	\$ 543,888

***Indirect Cost Information (To Be Completed by Your Business Office):**

If you are requesting reimbursement for indirect costs on line 10, please answer the following questions:

- (1) Do you have an Indirect Cost Rate Agreement approved by the Federal government? Yes No
- (2) If yes, please provide the following information:
 Period Covered by the Indirect Cost Rate Agreement: From: ___/___/___ To: ___/___/___ (mm/dd/yyyy)
 Approving Federal agency: ED Other (please specify): _____
- (3) For Restricted Rate Programs (check one) -- Are you using a restricted indirect cost rate that:
 Is included in your approved Indirect Cost Rate Agreement? or, Complies with 34 CFR 76.564(c)(2)?



U.S. DEPARTMENT OF EDUCATION
BUDGET INFORMATION
NON-CONSTRUCTION PROGRAMS

OMB Control Number: 1890-0004

Expiration Date: 06/30/2005

Name of Institution/Organization:
CRCL Inc.

Applicants requesting funding for only one year should complete the column under "Project Year 1." Applicants requesting funding for multi-year grants should complete all applicable columns. Please read all instructions before completing form.

SECTION B - BUDGET SUMMARY
NON-FEDERAL FUNDS

Budget Categories	Project Year 1(a)	Project Year 2 (b)	Project Year 3 (c)	Project Year 4 (d)	Project Year 5 (e)	Total (f)
1. Personnel	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
2. Fringe Benefits	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
3. Travel	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
4. Equipment	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
5. Supplies	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
6. Contractual	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
7. Construction	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
8. Other	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
9. Total Direct Costs (lines 1-8)	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
10. Indirect Costs	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
11. Training Stipends	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0
12. Total Costs (lines 9-11)	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0	\$ 0

ASSURANCES - NON-CONSTRUCTION PROGRAMS

OMB Approval No. 0348-0040

Public reporting burden for this collection of information is estimated to average 15 minutes per response, including time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding the burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Office of Management and Budget, Paperwork Reduction Project (0348-0040), Washington, DC 20503.

PLEASE DO NOT RETURN YOUR COMPLETED FORM TO THE OFFICE OF MANAGEMENT AND BUDGET. SEND IT TO THE ADDRESS PROVIDED BY THE SPONSORING AGENCY.

NOTE: Certain of these assurances may not be applicable to your project or program. If you have questions, please contact the awarding agency. Further, certain Federal awarding agencies may require applicants to certify to additional assurances. If such is the case, you will be notified.

As the duly authorized representative of the applicant, I certify that the applicant:

1. Has the legal authority to apply for Federal assistance and the institutional, managerial and financial capability (including funds sufficient to pay the non-Federal share of project cost) to ensure proper planning, management and completion of the project described in this application.
2. Will give the awarding agency, the Comptroller General of the United States and, if appropriate, the State, through any authorized representative, access to and the right to examine all records, books, papers, or documents related to the award; and will establish a proper accounting system in accordance with generally accepted accounting standards or agency directives.
3. Will establish safeguards to prohibit employees from using their positions for a purpose that constitutes or presents the appearance of personal or organizational conflict of interest, or personal gain.
4. Will initiate and complete the work within the applicable time frame after receipt of approval of the awarding agency.
5. Will comply with the Intergovernmental Personnel Act of 1970 (42 U.S.C. §§4728-4763) relating to prescribed standards for merit systems for programs funded under one of the 19 statutes or regulations specified in Appendix A of OPM's Standards for a Merit System of Personnel Administration (5 C.F.R. 900, Subpart F).
6. Will comply with all Federal statutes relating to nondiscrimination. These include but are not limited to: (a) Title VI of the Civil Rights Act of 1964 (P.L. 88-352) which prohibits discrimination on the basis of race, color or national origin; (b) Title IX of the Education Amendments of 1972, as amended (20 U.S.C. §§1681-1683, and 1685-1686), which prohibits discrimination on the basis of sex; (c) Section 504 of the Rehabilitation Act of 1973, as amended (29 U.S.C. §794), which prohibits discrimination on the basis of handicaps; (d) the Age Discrimination Act of 1975, as amended (42 U.S.C. §§6101-6107), which prohibits discrimination on the basis of age; (e) the Drug Abuse Office and Treatment Act of 1972 (P.L. 92-255), as amended, relating to nondiscrimination on the basis of drug abuse; (f) the Comprehensive Alcohol Abuse and Alcoholism Prevention, Treatment and Rehabilitation Act of 1970 (P.L. 91-616), as amended, relating to nondiscrimination on the basis of alcohol abuse or alcoholism; (g) §§523 and 527 of the Public Health Service Act of 1912 (42 U.S.C. §§290 dd-3 and 290 ee- 3), as amended, relating to confidentiality of alcohol and drug abuse patient records; (h) Title VIII of the Civil Rights Act of 1968 (42 U.S.C. §§3601 et seq.), as amended, relating to nondiscrimination in the sale, rental or financing of housing; (i) any other nondiscrimination provisions in the specific statute(s) under which application for Federal assistance is being made; and, (j) the requirements of any other nondiscrimination statute(s) which may apply to the application.
7. Will comply, or has already complied, with the requirements of Titles II and III of the Uniform Relocation Assistance and Real Property Acquisition Policies Act of 1970 (P.L. 91-646) which provide for fair and equitable treatment of persons displaced or whose property is acquired as a result of Federal or federally-assisted programs. These requirements apply to all interests in real property acquired for project purposes regardless of Federal participation in purchases.
8. Will comply, as applicable, with provisions of the Hatch Act (5 U.S.C. §§1501-1508 and 7324-7328) which limit the political activities of employees whose principal employment activities are funded in whole or in part with Federal funds.

Previous Edition Usable

Authorized for Local Reproduction

Standard Form 424B (Rev. 7-97)
Prescribed by OMB Circular A-102

Tracking Number: GRANT00057168

9. Will comply, as applicable, with the provisions of the Davis- Bacon Act (40 U.S.C. §§276a to 276a-7), the Copeland Act (40 U.S.C. §276c and 18 U.S.C. §874), and the Contract Work Hours and Safety Standards Act (40 U.S.C. §§327- 333), regarding labor standards for federally-assisted construction subagreements.
10. Will comply, if applicable, with flood insurance purchase requirements of Section 102(a) of the Flood Disaster Protection Act of 1973 (P.L. 93-234) which requires recipients in a special flood hazard area to participate in the program and to purchase flood insurance if the total cost of insurable construction and acquisition is \$10,000 or more.
11. Will comply with environmental standards which may be prescribed pursuant to the following: (a) institution of environmental quality control measures under the National Environmental Policy Act of 1969 (P.L. 91-190) and Executive Order (EO) 11514; (b) notification of violating facilities pursuant to EO 11738; (c) protection of wetlands pursuant to EO 11990; (d) evaluation of flood hazards in floodplains in accordance with EO 11988; (e) assurance of project consistency with the approved State management program developed under the Coastal Zone Management Act of 1972 (16 U.S.C. §§1451 et seq.); (f) conformity of Federal actions to State (Clean Air) Implementation Plans under Section 176(c) of the Clean Air Act of 1955, as amended (42 U.S.C. §§7401 et seq.); (g) protection of underground sources of drinking water under the Safe Drinking Water Act of 1974, as amended (P.L. 93-523); and, (h) protection of endangered species under the Endangered Species Act of 1973, as amended (P.L. 93- 205).
12. Will comply with the Wild and Scenic Rivers Act of 1968 (16 U.S.C. §§1271 et seq.) related to protecting components or potential components of the national wild and scenic rivers system.
13. Will assist the awarding agency in assuring compliance with Section 106 of the National Historic Preservation Act of 1966, as amended (16 U.S.C. §470), EO 11593 (identification and protection of historic properties), and the Archaeological and Historic Preservation Act of 1974 (16 U.S.C. §§469a-1 et seq.).
14. Will comply with P.L. 93-348 regarding the protection of human subjects involved in research, development, and related activities supported by this award of assistance.
15. Will comply with the Laboratory Animal Welfare Act of 1966 (P.L. 89-544, as amended, 7 U.S.C. §§2131 et seq.) pertaining to the care, handling, and treatment of warm blooded animals held for research, teaching, or other activities supported by this award of assistance.
16. Will comply with the Lead-Based Paint Poisoning Prevention Act (42 U.S.C. §§4801 et seq.) which prohibits the use of lead-based paint in construction or rehabilitation of residence structures.
17. Will cause to be performed the required financial and compliance audits in accordance with the Single Audit Act Amendments of 1996 and OMB Circular No. A-133, "Audits of States, Local Governments, and Non-Profit Organizations."
18. Will comply with all applicable requirements of all other Federal laws, executive orders, regulations, and policies governing this program.

* SIGNATURE OF AUTHORIZED CERTIFYING OFFICIAL Doug Cooper	* TITLE Executive Director
* APPLICANT ORGANIZATION CRCL Inc.	* DATE SUBMITTED 11-16-2005

Standard Form 424B (Rev. 7-97) Back

DISCLOSURE OF LOBBYING ACTIVITIES

Complete this form to disclose lobbying activities pursuant to 31 U.S.C. 1352
(See reverse for public burden disclosure.)

Approved by OMB

0348-0046

<p>1. * Type of Federal Action:</p> <p><input type="checkbox"/> a. contract</p> <p><input checked="" type="checkbox"/> b. grant</p> <p><input type="checkbox"/> c. cooperative agreement</p> <p><input type="checkbox"/> d. loan</p> <p><input type="checkbox"/> e. loan guarantee</p> <p><input type="checkbox"/> f. loan insurance</p>	<p>2. * Status of Federal Action:</p> <p><input checked="" type="checkbox"/> a. bid offer application</p> <p><input type="checkbox"/> b. initial award</p> <p><input type="checkbox"/> c. post-award</p>	<p>3. * Report Type:</p> <p><input checked="" type="checkbox"/> a. initial filing</p> <p><input type="checkbox"/> b. material change</p> <p>For Material Change Only:</p> <p>year quarter</p> <p>date of last report</p>
<p>4. * Name and Address of Reporting Entity:</p> <p><input checked="" type="checkbox"/> Prime <input type="checkbox"/> SubAwardee Tier if known:</p> <p>* Name: CRCL Inc.</p> <p>* Address: 820 Calle Pluma</p> <p>San Clemente CA 92673</p> <p>Congressional District, if known:</p>	<p>5. If Reporting Entity in No.4 is Subawardee, Enter Name and Address of Prime:</p>	
<p>6. * Federal Department/Agency:</p> <p>Department of Education</p>	<p>7. * Federal Program Name/Description: International Research and Studies</p> <p>CFDA Number, if applicable: 84.017</p>	
<p>8. Federal Action Number, if known:</p>	<p>9. Award Amount, if known:</p>	
<p>10. a. Name and Address of Lobbying Registrant (if individual, complete name):</p> <p>Name: none none</p> <p>Address:</p>	<p>b. Individual Performing Services (including address if different from No. 10a):</p> <p>Name: none none</p>	
<p>11. Information requested through this form is authorized by title 31 U.S.C. section 1352. This disclosure of lobbying activities is a material representation of fact upon which reliance was placed by the tier above when the transaction was made or entered into. This disclosure is required pursuant to 31 U.S.C. 1352. This information will be available for public inspection. Any person who fails to file the required disclosure shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.</p>	<p>Signature: Doug Cooper</p> <p>* Name: Mr. Doug Cooper</p> <p>Title: Executive Director</p> <p>Telephone No.: 949 496-5195</p> <p>Date: 11-16-2005</p>	
<p>Federal Use Only:</p>	<p style="text-align: right;">Authorized for Local Reproduction Standard Form - LLL (Rev. 7-97)</p>	

Public Burden Disclosure Statement

According to the Paperwork Reduction Act, as amended, no persons are required to respond to a collection of information unless it displays a valid OMB Control Number. The valid OMB control number for this information collection is OMB No. 0348-0046. Public reporting burden for this collection of information is estimated to average 10 minutes per response, including time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding the burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to the Office of Management and Budget, Paperwork Reduction Project (0348-0046), Washington, DC 20503.

NOTICE TO ALL APPLICANTS

The purpose of this enclosure is to inform you about a new provision in the Department of Education's General Education Provisions Act (GEPA) that applies to applicants for new grant awards under Department programs. This provision is Section 427 of GEPA, enacted as part of the Improving America's Schools Act of 1994 (Public Law (P.L.) 103-382).

To Whom Does This Provision Apply?

Section 427 of GEPA affects applicants for new grant awards under this program. **ALL APPLICANTS FOR NEW AWARDS MUST INCLUDE INFORMATION IN THEIR APPLICATIONS TO ADDRESS THIS NEW PROVISION IN ORDER TO RECEIVE FUNDING UNDER THIS PROGRAM**

. (If this program is a State-formula grant program, a State needs to provide this description only for projects or activities that it carries out with funds reserved for State-level uses. In addition, local school districts or other eligible applicants that apply to the State for funding need to provide this description in their applications to the State for funding. The State would be responsible for ensuring that the school district or other local entity has submitted a sufficient section 427 statement as described below.)

What Does This Provision Require?

Section 427 requires each applicant for funds (other than an individual person) to include in its application a description of the steps the applicant proposes to take to ensure equitable access to, and participation in, its Federally-assisted program for students, teachers, and other program beneficiaries with special needs. This provision allows applicants discretion in developing the required description. The statute highlights six types of barriers that can impede equitable access or participation: gender, race, national origin, color, disability, or age. Based on local circumstances, you should determine whether these or other barriers may prevent your students, teachers, etc. from such access or participation in, the Federally-funded project or activity. The description in your application of steps to be taken to overcome these barriers need not be lengthy; you may provide a clear and succinct

description of how you plan to address those barriers that are applicable to your circumstances. In addition, the information may be provided in a single narrative, or, if appropriate, may be discussed in connection with related topics in the application.

Section 427 is not intended to duplicate the requirements of civil rights statutes, but rather to ensure that, in designing their projects, applicants for Federal funds address equity concerns that may affect the ability of certain potential beneficiaries to fully participate in the project and to achieve to high standards. Consistent with program requirements and its approved application, an applicant may use the Federal funds awarded to it to eliminate barriers it identifies.

What are Examples of How an Applicant Might Satisfy the Requirement of This Provision?

The following examples may help illustrate how an applicant may comply with Section 427.

- (1) An applicant that proposes to carry out an adult literacy project serving, among others, adults with limited English proficiency, might describe in its application how it intends to distribute a brochure about the proposed project to such potential participants in their native language.
- (2) An applicant that proposes to develop instructional materials for classroom use might describe how it will make the materials available on audio tape or in braille for students who are blind.
- (3) An applicant that proposes to carry out a model science program for secondary students and is concerned that girls may be less likely than boys to enroll in the course, might indicate how it intends to conduct "outreach" efforts to girls, to encourage their enrollment.

We recognize that many applicants may already be implementing effective steps to ensure equity of access and participation in their grant programs, and we appreciate your cooperation in responding to the requirements of this provision.

Estimated Burden Statement for GEPA Requirements

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless such collection displays a valid OMB control number. The valid OMB control number for this information collection is 1890-0007. The time required to complete this information collection is estimated to average 1.5 hours per response, including the time to review instructions, search existing data resources, gather the data needed, and complete and review the information collection. **If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to:** Director, Grants Policy and Oversight Staff, U.S. Department of Education, 400 Maryland Avenue, SW (Room 3652, GSA Regional Office Building No. 3), Washington, DC 20202-4248.

Attachment Information

File Name	Mime Type
1342-gepa427.pdf	application/pdf

Compliance with The U.S. Department of Education's General Education Provisions Act (GEPA) Section 427

We welcome the challenge of seeking innovative ways to ensure compliance with the U.S. Department of Education's General Education Provisions Act (GEPA) Section 427. Because we are preparing instructional materials, we follow example 2 of the guidelines, and describe our approach to ensuring equitable access to, and use of, SEAlang Lab resources.

1. We are committed to using technology to make our resources as accessible as possible by the disabled. In designing the SEAlang Lab user interfaces, we will make every effort to comply with the **Web Accessibility Initiative**, which addresses the needs of those with disabilities.
2. The tools will be universally and freely accessible via the Internet, ensuring equitable access to all facilities.
3. The tools themselves are designed to accommodate as wide a variety of user and learning styles as possible, include meeting the needs of students with perceptual or physical problems that might interfere with the reading of complex-script languages. We will allow choice in, and modification of, font and text display styles.
4. Although the SEAlang Lab cannot fully accommodate blind users, the pre-recorded audio lexicon we link to the tools will be of great assistance to some blind and partially blind users. We will attempt to cooperate with the Thai Association for the Blind in making their DAISY audio-texts available through the SEAlang Lab facilities.
5. The SEAlang Lab is meant to be an open, extensible resources. We will encourage Southeast Asian Open Source software projects we work with to use SEAlang Lab tools and data in creating additional **assistive technology** resources, and provide any help that we can.

Certification Regarding Debarment, Suspension, Ineligibility and Voluntary Exclusion - Lower Tier Covered Transactions

This certification is required by the Department of Education regulations implementing Executive Order 12549, Debarment and Suspension, 34 CFR Part 85, for all lower tier transactions meeting the threshold and tier requirements stated at Section 85.110.

Instructions for Certification

1. By signing and submitting this proposal, the prospective lower tier participant is providing the certification set out below.
2. The certification in this clause is a material representation of fact upon which reliance was placed when this transaction was entered into. If it is later determined that the prospective lower tier participant knowingly rendered an erroneous certification, in addition to other remedies available to the Federal Government, the department or agency with which this transaction originated may pursue available remedies, including suspension and/or debarment.
3. The prospective lower tier participant shall provide immediate written notice to the person to which this proposal is submitted if at any time the prospective lower tier participant learns that its certification was erroneous when submitted or has become erroneous by reason of changed circumstances.
4. The terms "covered transaction," "debarred," "suspended," "ineligible," "lower tier covered transaction," "participant," "person," "primary covered transaction," "principal," "proposal," and "voluntarily excluded," as used in this clause, have the meanings set out in the Definitions and Coverage sections of rules implementing Executive Order 12549. You may contact the person to which this proposal is submitted for assistance in obtaining a copy of those regulations.
5. The prospective lower tier participant agrees by submitting this proposal that, should the proposed covered transaction be entered into, it shall not knowingly enter into any lower tier covered transaction with a person who is debarred, suspended, declared ineligible, or voluntarily excluded from participation in this covered transaction, unless authorized by the department or agency with which this transaction originated.
6. The prospective lower tier participant further agrees by submitting this proposal that it will include the clause titled ACertification Regarding Debarment, Suspension, Ineligibility, and Voluntary Exclusion-Lower Tier Covered Transactions, and without modification, in all lower tier covered transactions and in all solicitations for lower tier covered transactions.
7. A participant in a covered transaction may rely upon a certification of a prospective participant in a lower tier covered transaction that it is not debarred, suspended, ineligible, or voluntarily excluded from the covered transaction, unless it knows that the certification is erroneous. A participant may decide the method and frequency by which it determines the eligibility of its principals. Each participant may but is not required to, check the Nonprocurement List.
8. Nothing contained in the foregoing shall be construed to require establishment of a system of records in order to render in good faith the certification required by this clause. The knowledge and information of a participant is not required to exceed that which is normally possessed by a prudent person in the ordinary course of business dealings.
9. Except for transactions authorized under paragraph 5 of these instructions, if a participant in a covered transaction knowingly enters into a lower tier covered transaction with a person who is suspended, debarred, ineligible, or voluntarily excluded from participation in this transaction, in addition to other remedies available to the Federal Government, the department or agency with which this transaction originated may pursue available remedies, including suspension and/or debarment.

Certification

- (1) The prospective lower tier participant certifies, by submission of this proposal, that neither it nor its principals are presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded from participation in this transaction by any Federal department or agency.
- (2) Where the prospective lower tier participant is unable to certify to any of the statements in this certification, such prospective participant shall attach an explanation to this proposal.

* NAME OF APPLICANT

CRCL Inc.

* PRINTED NAME AND TITLE OF AUTHORIZED REPRESENTATIVE

* Name: Doug Cooper

* Title: Executive Director

Signature: Doug Cooper

Submitted Date: 11-16-2005

CERTIFICATION REGARDING LOBBYING

Certification for Contracts, Grants, Loans, and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No Federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of an agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any Federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers (including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements) and that all subrecipients shall certify and disclose accordingly. This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

Statement for Loan Guarantees and Loan Insurance

The undersigned states, to the best of his or her knowledge and belief, that:

If any funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this commitment providing for the United States to insure or guarantee a loan, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions. Submission of this statement is a prerequisite for making or entering into this transaction imposed by section 1352, title 31, U.S. Code. Any person who fails to file the required statement shall be subject to a civil penalty of not less than \$10,000 and not more than \$100,000 for each such failure.

* APPLICANT'S ORGANIZATION CRCL Inc.
* PRINTED NAME AND TITLE OF AUTHORIZED REPRESENTATIVE Prefix: * First Name: Doug Middle Name: * Last Name: Cooper Suffix: * Title: Executive Director
* SIGNATURE: Doug Cooper * DATE: 11-16-2005

Project Narrative

Abstract Narrative

Attachment 1:

Title: Pages: Uploaded File: **7018-irs05-abstract.pdf**

The SEALang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition in Complex-Script Languages

Prepared learning resources for less commonly taught languages (LCTLs) are rarely sufficient. Nevertheless, broad reading of **authentic texts** usually helps mitigate the lack of ancillary study or reference materials – and even, for some self-directed students, of teachers or native speakers.

But dozens of languages in Southeast, Central, and South Asia and the Middle East (including Arabic, Thai, and Urdu) require **complex scripts**. They use non-roman alphabets, non-linear ordering, context-dependent letterforms or ligatures, implicit vowels, tones, and registers, and may not even space between words. Students who turn to authentic materials are deadlocked: they must read extensively to acquire vocabulary, yet cannot until they already *have* vocabulary.

The **SEALang Lab** takes a new approach to **reading, writing, and vocabulary acquisition** for complex-script LCTLs. Our goals are well-defined and achievable; we will build:

- the **Reader's Workbench**: a free, online *complex-script reading tool* that increases reading speed and accuracy at all levels, using any text (such as Web-accessible newspapers). Besides integrating dictionary and corpus data, the Workbench provides word and phrase segmentation, automates phonetic transcription, highlights core-vocabulary coverage, evaluates text difficulty, and even finds appropriately graded texts for the student.
- the **Writer's Workbench**: a free, online *complex-script writing tool*. The tool addresses both mechanical ability (using predictive completion from local script, phonetic transcription, or letter-by-letter transliteration), and syntactic / expressive competence (via integrated corpus and collocational reference tools).
- the **Vocabulary Workbench**: a free, online *data-driven drill and test tool*. Using novel heuristics that draw on dictionary and corpus data, the Workbench dynamically generates review and reinforcement based on student queries, arbitrary texts, or required word lists. It can be coupled to the Reader's Workbench, or used standalone as a teacher's assistant.

The Workbenches are enabled by enriched lexical resources derived from the USED/CRCL-sponsored **SEALang Library** by extending the functionality of e-text corpora, and adding word-origin, frequency, difficulty, collocational, usage, and core-wordlist data to existing dictionaries.

The SEALang Lab proposal is submitted by the **Center for Research in Computational Linguistics** (CRCL Inc., a US 501(c)3 nonprofit). Its design has benefited greatly from wide, public demonstration of our preliminary development, most recently as the invited plenary speaker at the September 2005 **Interagency Language Roundtable** in Washington, DC. All tools will be class-tested and improved in collaboration with the nation's best intensive language programs: the **Foreign Service Institute**, the **Defense Language Institute**, U Wisconsin-Madison's **Center for Southeast Asian Studies** and the **Southeast Asian Studies Summer Institute**.

Our working language is Thai. Unlike Arabic or Urdu, Thai has all of the characteristic complex script problems, including lack of word separation. Southeast Asia's strategic importance to the US makes Thai very attractive as a cross-training source or target *vis a vis* Khmer, Burmese, and Lao, as well as Shan and other ethnic minority languages. Finally, despite its difficulty, Thai has sufficient existing resources and students for timely, cost-effective development and testing.

In response to the 2005 program priority for the languages of Islamic societies, we will also target (**Pattani**) **Malay** and its **Arabic-based Jawi** script. Universally taught in troubled Southern Thailand's 500-plus Pondok schools, Jawi typifies Arabic scripts used to write non-Arabic languages (Urdu, Pashto, etc), and is rapidly regaining its historical popularity in Southeast Asia.

Project Narrative

Project Narrative

Attachment 1:

Title: Pages: Uploaded File: **2806-Mandatory_irs05.pdf**

Attachment 2:

Title: Pages: Uploaded File: **203-irs05-contents.pdf**

1. NEED FOR THE PROJECT

Introducing the SEALang Lab The Center for Research in Computational Linguistics (**CRCL Inc.**), working in collaboration with the Foreign Service Institute (**FSI**), the Defense Language Institute (**DLI**), the Southeast Asian Studies Summer Institute (**SEASSI**), and the University of Wisconsin-Madison Center for Southeast Asian Studies (**CSEAS**), proposes:

- to develop **assistive technology** tools that dramatically reduce problems of reading, writing, and studying languages that require complex scripts, and to make them freely available as on-line **software workbenches** (see examples in Appendices 1a/1b),
- to pioneer new approaches to **dynamically generating** drill, test, and review, with or without teacher assistance, to support vocabulary acquisition in resource-poor, less commonly taught languages (see examples in Appendix 1c),
- to help teaching faculty pursue badly needed research into **complex-script pedagogy**, and test alternative approaches to reading, writing, and vocabulary acquisition.

Our demonstration language is **Thai**, a key regional language that presents all of the problems, including complex orthography, and lack of word segmentation, found in a very broad range of complex scripts that are used throughout Asia and the Middle East.

We will also deliver tools for **Pattani Malay**, and its alternative **Arabic-based Jawi**, non-segmented Thai, and romanized Rumi scripts. Jawi script extends Arabic; it is found in Islamic communities across Southeast Asia, and is universally taught in Southern Thailand's 500-plus Pondok schools. Supported by a new state bilingual education policy, it is rapidly gaining popularity in Thailand's disturbed Muslim-majority southernmost provinces (see also Appendix 5).

The SEALang Lab's assistive technology tools will support traditional **classroom teaching** and independent **autonomous study** at all stages. They will help language professionals raise their reading and writing abilities to match speaking and comprehension, and will enable **cross-training** – teaching speakers of related or heritage languages to read, write, or speak as rapidly as possible. Finally, they will help researchers raise, test, and formally evaluate longstanding methodological questions about the best ways to teach and learn complex-script languages.

1a) Need We need the SEALang Lab project because writing systems throughout Asia and the Middle East – आप से मिलकर बहुत खुशी हुई, ལེན ལྟོ འཕྲུལ འཕྲུལ།, تشرّفنا , ยินดีที่ได้รู้จัก, ஸ்ரீயவில்லை, and dozens more – are hard in a way that Western writing systems (even **очень приятно** and **χάρηκα για τη γνωριμία**) are not. Western languages like Russian and Greek may use non-roman *alphabets*, but most Asian and Middle Eastern systems use **complex scripts**.

Building America’s capacity to teach, learn, and use complex-script languages requires a new generation of reading, writing, and study tools. We call them **workbenches**: they are intuitive, user-friendly workplaces that keep tools and reference materials close at hand. All SEALang Lab workbenches can be used on-line, and do not require downloads or locally installed software.

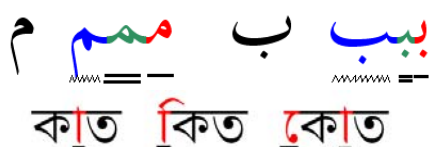
- The **Reader’s Workbench** (see Appendix 1a) enables fast, accurate, close *intensive* or broad *extensive* reading and navigation of any text or Web page. Functions include:
 - *transcribing* the text into IPA phonetics or alternative (Jawi / Rumi / Thai) scripts,
 - *segmenting* non-spaced complex scripts into words, compounds, or longer phrases,
 - *estimating difficulty* of texts, and evaluating core-vocabulary coverage,
 - *seeking* texts with particular kinds of vocabulary content or levels of difficulty,
 - *immediate or delayed testing* via the Vocabulary Workbench, described below,
 - *integrating* all of the reference facilities provided by the SEALang Library (see below), including mono- and bilingual dictionaries, and monolingual and aligned-bitext corpora.
- The **Writer’s Workbench** (see Appendix 1b) helps students acquire the syntactic / expressive competence required for effective writing, and assists typing and spelling. It features:
 - *entry* using native script or phonetic transcription, with optional ‘smart’ approximation,
 - *predictive completion* of entries, sorted by corpus-based likelihood or word length,
 - *intelligent approximation* in attempting to complete misspelled words,
 - *predictive phrasal completion* of multi-word units and collocates,
 - *reverse lookup* via English definitions; particularly helpful for technical terms,
 - *integrated reference* to allow corpus-based comparison of words or phrases.

- The **Vocabulary Workbench** (see Appendix 1c) automatically devises, delivers, and tracks drill and review. It can be used standalone, or in a Reader's Workbench session. It can:
 - *dynamically generate* a wide variety of drill and review questions from lexical data,
 - use *text, sound, typed input*, and *timers* to build both receptive and productive drills
 - *progressively cover core vocabularies*, and base drills on reading or required word lists,
 - *record* user performance, and track the effectiveness of different test types,
 - *assist the instructor* by generating tests that can be modified and saved for later use.

Important parts of SEAlang Lab functionality will be based on existing lexical resources supplied by the **SEAlang Library**, a four-year project funded by the USED *Technological Innovation for Cooperation on Foreign Information Access* (TICFIA) program (see Appendices 3 and 4). We will extend these resources to increase Workbench capabilities, and provide a 24,000-entry Thai audio lexicon to support vocabulary drill and reading skills development.

Why do scripts make some languages hard? Three features – *complex scripts*, *complex orthography*, and *non-segmentation* – make it extremely difficult to develop true, native-level **automaticity** in reading: the ability to instantly recognize a word's boundaries, sound, and meaning.

A **complex script** has a variety of unusual characteristics, including a non-roman alphabet, non-linear ordering, and context-dependent letterforms and ligatures (combining letters).



Four forms of the Arabic letters 'beh' and 'meem' showing initial (red), medial (green), final (blue), and isolate (black) letterforms.

Bengali vowels in red: **ka**, **ki**, **ko**

Fig 1. Arabic reads right-to-left, and Indic-derived scripts (including Thai, Burmese, and Khmer) are syllable **clusters**. (examples Sharon Correll; underlines added for non-color display)

The unwritten parts of a complex script are just as important. Tones, vowels, and other features language learners need to know are often not shown at all:

כל בני האדם נולדו בני חורין ושווים בערכם ובזכויותיהם.
כל בני האדם נולדו בני חורין ושווים בערכם ובזכויותיהם.

Fig 2. The upper text is written conventionally; the lower text shows all vowels. (text Omniglot)

<u>เก</u>	<u>แ</u> เก	ก <u>ุ</u>	ก <u>า</u>	ก <u>ิ</u>	ก <u>อ</u>	<u>ไ</u> ก	<u>เก</u> อ	<u>เก</u> าะ	<u>เก</u> ีย	<u>เก</u> ียะ	ข <u>น</u> ม
kee	kɛɛ	ku	kaa	kii	koo	kay	kəə	kɔʔ	kia	kiaʔ	kʰ <u>a</u> nom

Fig 3. Some (not all) Thai vowel forms in *blue* (over- and under-vowels are not underlined). One vowel may have as many as four parts orbiting the consonant. The final word has no written vowels; *a* and *o*, as well as tone (not shown), are implicit in the consonants.

A **complex orthography** has an unpredictable relationship between spelling and sound.

Spanish and Malaysian have simple orthographies in which spelling follows sound. But conservative writing systems like English and Thai have complex orthographies: spelling is heavily influenced by each word's etymological derivation. We write 'station' (not the phonetic 'stahshun'), and Thais write the cognate term **สถาน** (not phonetic **ชะทาน** /satʰaan/) because both words share a common Indo-European root, preserved in pre-Sanskrit Vedic as sthāna.

พัน พรรค์ พรรณ พันธ พันธุ์ ภณฑ์ พัญจน์

Fig 4. Pali-Sanskrit-influence on Thai words pronounced identically as pʰan. Only the first is spelled regularly; the rest preserve etymological detail. Khmer and English loans are similar.

Finally, some scripts are **unsegmented** – they are written continuously, without spaces between words. Although native speakers manage the process intuitively, accurately splitting unsegmented texts into individual words or compounds is a persistent source of problems even for relatively advanced language learners (as well as an extremely difficult computational task).

ลำ นำ นัน ไหล มา จาก ทิศ ตะวัน ตก ทอด ตัว เลอย เลอย... *segmented into words*
 ลำนำ นัน ไหลมา จาก ทิศตะวันตก ทอดตัว เลอยเลอย... *segmented into compounds*
 ลำนำนันไหลมาจากทิศตะวันตก ทอดตัวเลอยเลอย... *printed conventionally*

Fig 5. Thai (shown here), Burmese, Khmer, Lao, and other SEA scripts are all unsegmented.

In conventional segmented scripts, momentarily skipping past an unknown word to try to guess its meaning from context is a standard reading strategy. But in unsegmented scripts, lack of automaticity makes it very difficult to instantly see where the next whole word or compound begins. It is very easy to lose one's place, or to become confused by trying to make sense of just one part of a compound word, and then have to backtrack to the very beginning of a sentence.

Languages that use complex scripts, have complex orthography, and are unsegmented are invariably **less commonly taught languages** (LCTLs) as well. While nearly all LCTLs lack sufficient teachers and ancillary materials, students of roman-script LCTLs can usually fall back on reading to mitigate these shortfalls. It isn't uncommon to hear people speak of acquiring a reading knowledge of a second or third European language.

But complex scripts and orthographies (and, in Southeast Asia, lack of spacing) make it extraordinarily difficult to read one's way to competence in Thai, Arabic, and dozens of similar LCTLs. They discourage *writing*, which becomes increasingly important in attaining intermediate and advanced ability. Every aspect of learning suffers, from vocabulary acquisition to mastery of grammar and usage; indeed, the FSI ranks such languages as more or exceptionally difficult to learn (see <http://www.nvta.gov/lotw/months/november/learningExpectations.html>).

1b) National significance Complex scripts are found across all parts of Asia and the Middle East. Languages that use them are of increasing political and economic importance to the United States, as is recognition of our need to develop expertise in a wider range of world languages.

The SEALang Lab addresses issues of complex-script languages in general, and of two of the most difficult – Thai and Jawi – in particular. The Lab's priority for and significance to the national interest are readily shown by our collaborators and supporters, and our choice of demonstration languages is backed up by the importance of Southeast Asia.

The project's value to those whose careers are dedicated to the national interest is seen in commitments by the Foreign Service Institute (**FSI**) and the Defense Language Institute (**DLI**) to collaborate in class-testing the SEALang Lab, and in an enthusiastic assessment by the Coordinator of the Interagency Language Roundtable (**ILR**¹) of the quality and potential applications of our preliminary work. See Appendices 10 and 11 for all letters of collaboration and support.

Further collaboration in class-testing by the **CSEAS** and **SEASSI** language programs, and endorsements from the National Council of Less Commonly Taught Languages (**NCOLCTL**),

¹ This forum (<http://www.govtilr.org>) presents language-related technology to representatives of more than 30 Federal agencies, including the Departments of Education, State, and Defense.

SEASSI, the Council of Teachers of Southeast Asian Languages (**COTSEAL**), and the Linguistic Data Consortium (**LDC**) show our national significance in America's LCTL community: the SEALang Lab addresses widely recognized problems of complex-script language acquisition.

Our choices of Thai and the Arabic-based Jawi script that can be used for Standard and Pattani Malay are also based on need. Current student numbers for these languages are not the only issue; American preparedness and capacity to engage the region are also relevant concerns.

Thai is one of 14 **stronghold languages** designated by the DoD *2015 Strategic Language Requirements* (Malay is also named; five of the 28 languages are Southeast Asian). Arguably mainland SEA's lingua franca due to Thailand's heavyweight economic influence, Thai has 65 million native speakers, and 75 related dialects spoken from Southern China to Northern Malaysia, and from India's Assam state across Burma and Laos to the highlands of North Vietnam.

Thailand is strategically important to the U.S. It is formally designated as a major non-NATO ally, and is a critical transit point for our military operations in the Middle East and Central Asia. The annual **Cobra Gold** joint Armed Forces exercise, held in Thailand to improve combat readiness and interoperability, is one of the largest U.S. Asia-Pacific exercises. The Navy's bilateral **CARAT** exercises play a similar role – after all, central SEA is also a critical shipping route: 30% of the world's trade goods pass through Malaysia's Malacca Straits, including half its oil (and most of the oil used by Japan and South Korea, our strongest East Asian allies).

The ongoing Islamic insurgency in Southern Thailand – which is creating severe internal stress and is disrupting Thailand's relations with bordering Muslim-majority Malaysia – makes this trade particularly significant. In an unusually pointed statement in Bangkok, U.S. Ambassador to Thailand Ralph Boyce was reported saying that 'the United States is worried the violence in Thailand's deep South could develop into full-scale international terrorism' (**Bangkok Post**, 10/14/05). We discuss this situation further in Section 13, *Invitational Priorities*.

2. USEFULNESS OF THE PRODUCT TO OTHERS

In this section, we discuss the contribution of the SEALang Lab to Southeast Asian language

learning. We look at the present needs of existing constituencies, and describe the relatively new application of language cross-training. Section 5, *Contribution to Other Programs*, discusses the broader impact of our work on complex-script and LCTL pedagogy in general.

The need for tools is clear. As Prof. Teresita Ramos recently noted, development of resources for teaching SEA languages has been ‘excruciatingly slow compared to, say, the development of the so-called “other” less commonly taught languages such as Chinese, Japanese, and Korean’ (*The Politics of Teaching Southeast Asian Languages in the United States*, SEALS XIV 2004).

We have surveyed the potential SEALang Lab audience carefully: in meetings with students and faculty, in discussions on the SEALteach listserv, in presentations to (and follow-up discussions with) the COTSEAL and SEASSI boards at the 2005 AAS conference, to the ILR July Showcase, to the full ILR in September (as invited plenary speaker), to the Consortium for Advanced Study of Thai (CAST) in August, to the Center for Advanced Study of Language (CASL) in October, and to other agencies, centers, and groups. We have not one, but many audiences:

- First are programs that are collaborating in class-testing and improving our tools: the **full-year intensive** programs of the DLI and FSI, the **full-year regular** programs of U Wisconsin’s CSEAS, and the **intensive** and **heritage** programs of SEASSI. Our nation’s best, most influential language programs are unanimous in recognizing the SEALang Lab’s usefulness.
- Growing use of distance education and self-directed instruction has led to a new breed of **autonomous learners**; they shoulder the responsibilities traditionally taken by the instructor. Because prepared online curricula are generally unavailable for LCTLs, these students must locate texts of appropriate difficulty for study, prepare drill and test materials, and evaluate their own progress. The SEALang Lab assists in all these tasks.
- **Teaching faculty** face increasing demands to develop competence in, and teaching materials for, new languages and student constituencies. Yet we have little experimental evidence about what works best in meeting new needs, including teaching heritage learners, cross-training to related languages, and reaching out to K-12 populations that are now beginning to study LCTLs. The SEALang Lab gives teachers the instrumentation needed to test new ideas

for modernizing traditional methods, and facilitates research and professional development.

- **Language professionals**, including analysts, the foreign service corps, and military personnel, also have a stake in assistive technology for language learning and use. New initiatives, like the State Department’s *Language Learning Continuum* and the DoD *Language Transformation Roadmap*, make a career-long commitment to acquiring and improving language skills a basic requirement for service and advancement. The SEAlang Lab directly addresses the needs of this demanding, highly motivated population of adult learners.

Cross-Training Cross-training uses language similarity to speed learning. It is of growing importance as American policy-makers recognize the need for broad language competence and true regional expertise. Cross-training is especially helpful for LCTLs (like the Arabic-script, Indo-European languages Pashto and Urdu) whose unfamiliar scripts conceal loan or cognate words. It is also of critical value in enabling *surge response* to emergency needs, particularly for ethnic minority languages that are related to better-known national languages.

The SEAlang Lab will make an immediate contribution to language cross-training for South-east Asian languages. These share many relations: some **genetic** (like the Tai-Kadai languages Thai / Lao / Shan, or the Mon-Khmer languages Khmer / Vietnamese / Wa), some **contact** (like Thai / Khmer), and some through common **loans** (as in the Pali-Sanskrit vocabulary Thai shares with Burmese, Malay, and the other SEA languages, or with English via Indo-European). One-third of written Thai is probably Pali-Sanskrit in origin, and 20% or more may be Khmer loans.

Thai	มนุษย์ทั้งหลายเกิดมามีอิสระและเสมอภาคกันในเกียรติศักดิ์
Khmer	មនុស្សទាំងឡាយកើតមកមានសេរីភាពនិងភាពស្មើ ៗ គ្នាក្នុងសិទ្ធិនិងសេចក្តីថ្លៃថ្នូរ ។
Lao	ມະນຸດທຸກຄົນເກີດມາມີກຽດສັກສີ ສິດທິ ເສລີພາບແລະຄວາມສເມີພາບເທົ່າທຽມກັນ.
Arabic (for Jawi)	سموا <u>مأنوسيا</u> دلاحيركن بييس دان سما راتا داري ساكي كموليان دان حق-حق.
Malay	Semua <u>manusia</u> dilahirkan bebas dan samarata dari segi kemuliaan dan hak-hak.

Fig. 6. ‘Humans are born ...’ The common Indic term for ‘human’ is blue in all languages. The first Thai/Lao-Khmer cognates are yellow, with Khmer ‘born’ double-lined. Thai shows additional Indic in green (other languages have similar Indic influence). An Arabic font approximates Jawi, which has 6 additional letters and awaits Unicode approval (Appendix 5b).

Studying a European language without looking for cognates from Greek or Latin roots seems inconceivable; even native English speakers get this preparation (in studying for the SAT) to help them read college English. But SEA language teaching regularly ignores the equivalent Pali-Sanskrit connection, as well as other regional genetic and contact links.

The SEAlang Lab specifically addresses cross-training. We will extend the SEAlang Library's Thai-English dictionary to mark Pali-Sanskrit and Mon-Khmer words (drawing on published resources and our own prior work), and to recognize so-called *word families* that have been borrowed wholesale. The Reader's Workbench will highlight these words, and the Vocabulary Workbench will use the same information to create quizzes and drills.

The real payoff, though, is not for learning Thai, but for *using* Thai to learn Khmer, Lao, or related languages. For example, Wa (Mon-Khmer), Karen (Tibeto-Burman) and Shan (Tai-Kadai) are spoken throughout Northern Burma and across its borders in Thailand, China, Laos, and India. This is a highly volatile region, and these languages may easily be tomorrow's Ache-nese or Chechen. None have substantial teaching resources, and cross-training from Thai, Khmer, or Burmese is by far the fastest route to language competence should the need arise.

3. ACCOUNT OF RELATED MATERIALS

3a) Citation of materials Thai has some basic segmentation utilities; e.g. SWATH <http://www.cs.cmu.edu/~paisarn/software.html> and ICU <http://www-306.ibm.com/software/globalization/topics/thaiusabilities/text.jsp> are used to embed line-breaks in Thai Web pages. Our own Thai Internet Education (TIE) project <http://crcl.th.net/index.htm> had a rudimentary implementation of glossed, machine-segmented text in 1996 in its *Thai On-Line Library*. The few commercial tools (*Rosetta Stone*, *Learning Thai Script CD-Rom*) are specialized and limited.

The TICFIA-funded *SEAlang Library* <http://crcl.th.net/sealang> project is preparing ten bilingual dictionaries for seven complex-script Southeast Asian languages (Thai, Burmese, Khmer, Lao, Mon, Shan, and Karen), and Vietnamese. The Library will provide dictionary search tools, extensive monolingual and aligned bitext corpora, and a variety of software tools for approxi-

mate search, automatic generation of Library of Congress romanization, as well as a simple reader's helper and text-as-image delivery service. See also Appendices 3 and 4.

The IRS-funded *Teaching and Learning Aids for Linguistically Complex Languages* (2006–2009) project at the Linguistic Data Consortium (LDC) <http://www ldc upenn edu> is developing a set of tools for dictionary lookup, reading facilitation, word recognition assessment, and text annotation for use with languages that have *complex morphology* or word structure. They are implementing tools for Arabic and Nahuatl. The project's primary focus is morphological parsing to facilitate lookup of inflectional forms in both dictionaries and text corpora.

The IRS-funded *Web-Based Advanced Translation Course for Thai, Indonesian, and Tagalog* (2003–2006) at Northern Illinois University <http://www.seasite.niu.edu/trans/Thai> focuses on the theory and practice of translation, and provides some tools for word frequency and text concordancing. An interlinear translation aid helps organize text, but does not assist in translation. The NIU *SEAsite Language & Culture* pages <http://www.seasite.niu.edu/> also provide some Thai materials, usually for their introductory courses. A Thai-English dictionary (8,000 entries; many are literary citations or are borrowed from So Sethabutr's work) is its largest resource.

Some general-purpose Web-accessible tools link Web pages to dictionaries; these include *Wordbot* <http://www.wordbot.com> and *Gymn@zilla* <http://140.127.211.213/cgi-bin/gz-cgi/-gzstart.pl>. These are only effective with segmented writing systems, and make only rudimentary attempts to find citation forms of derived or inflected entries. Other tools rely heavily on text corpora, and are typically developed under the *data-driven learning* rubric. Mike Scott's *Word-Smith Tools* <http://www.lexically.net/wordsmith> are quite widely used, but cannot handle non-segmented texts and have problems with some complex scripts. Adam Kilgarriff's 'word sketch' system *WASPS* <http://wasps.itri.bton.ac.uk> attempts to analyze and exemplify English word senses by locating characteristic collocational patterns; it is mainly conceived as a lexicographer's assistant, but is very useful for students as well.

Tom Cobb's *Compleat Lexical Tutor* at www.lextutor.ca is a data-driven system; it focuses on French / English acquisition, and provides a variety of drills and 'drill-builder' tools (e.g. it can

generate cloze tests, extract corpus examples, and even has a ‘reaction time instrument builder’ to help test automaticity). Its on-line vocabulary test (based on Nation and Laufer) covers 1,000 through 5,000-word levels, an academic wordlist and a 10,000-word master list.

The REAP system <http://orleans.lti.cs.cmu.edu/Reap/> is designed for native-speaker English learners, but provides some interesting functionality. The system helps *a)* assess text difficulty, *b)* locate texts that have a particular difficulty level, and *c)* assess student ability with such texts. REAP takes a lexically oriented approach to measuring text difficulty (in contrast to the traditional sentence-oriented measures). REAP also automatically generates a half-dozen vocabulary question types using English WordNet dictionary data; see <http://wordnet.princeton.edu>.

Tools that assist key entry by predictive completion are widely studied for applications that help the disabled and, more recently, to enable faster text messaging (e.g. the T9 standard). Preliminary work has begun for Thai, but there are no standards or available test data sets.

Finally, research on script-related questions in language acquisition is rare. Red² discusses reading complex scripts in a fundamental paper that defines many of the issues that need to be considered, but which were never pursued. Fukkink³ also raises one important question we wish to pursue – will skills training at the word or phrase level improve automaticity in text reading? – but only looks at a single Roman script. Webb⁴ compares receptive to productive learning of vocabulary, noting that ‘to my knowledge, there has been no research comparing receptive and productive learning’ other than indirect studies involving word pairs.

There is much work on vocabulary acquisition, but not for complex-script languages per se. Nation⁵ provides an excellent text on the subject, with particular attention to core vocabularies

² Red, David L. 1999. *Adults learning to read a second script: What we’ve learned*. 50th Georgetown University Round Table on Languages and Linguistics 1999, GU Press, 2001

³ Fukkink, RG, et al 2005 *Does Training in Second-Language Word Recognition Skills Affect Reading Comprehension?*. *Modern Language Journal*, 89 (1).

⁴ Webb, S. 2005. *Receptive and productive vocabulary learning: effects of reading and writing on word knowledge*. *Studies in Second Language Acquisition*, Cambridge U Press 27 (1). 33-52.

⁵ Nation, Paul, 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.

and text comprehension. Laufer⁶ raises critical questions about the effectiveness of extensive reading in vocabulary acquisition, and the importance of cognitive effort in retention. Hill⁷ takes the next step into experimental research using digital resources.

3b) Commentary As a rule, Thai-specific resources do not address general issues of reading, writing, and vocabulary acquisition. While there are both dictionary and corpus reference resources, they are not integrated into reading, writing, or study tools. The NIU word frequency and concordance tools require hand-segmented texts prepared for NIU courses.

General-purpose tools are not readily adapted to Thai and similar complex-script languages. Text display facilities like Gymn@zilla and REAP are not designed to deal with nonsegmented texts (indeed, even sophisticated search engines like Google will only index and locate texts that have been segmented in advance). The few drill-generating tools, like the Compleat Lexical Tutor or REAP, either use prepared test material or rely on English WordNet data.

Overall, language-specific tools (except for very widely taught languages) are just beginning to address a small subset of problems, while general-purpose tools are not easily adapted to languages like Thai. In some cases, algorithms and ideas can be adapted; e.g. techniques used by REAP to assess text difficulty by lexical content, rather than via analysis of sentence structure.

3c) Duplication The SEALang Lab will not duplicate other materials. The Lab is designed to complement the SEALang Library; we provide tools that use the Library's data resources, and will extend its Thai-English dictionary data. We also extend the NIU *Thai Translation Course* resources, which focus on NIU course content; we are producing general-purpose tools.

The LDC's plan for complex morphology tools also has some similarity to ours, but to the extent they intersect, our plans are completely complementary. Their interest is in languages with complex morphology (even when simple scripts are involved, as in the case of Turkish),

⁶ Laufer, Batia, 1997, *The lexical plight in second language reading: words you don't know, words you think you know and words you can't guess*. In *Second Language Vocabulary Acquisition*, eds. J. Coady and T. Huckin. Cambridge University Press. pp. 20-34.

⁷ Hill, Monica, and Laufer, Batia. 2003 *Type of task, time-on-task, and electronic dictionaries in incidental vocabulary acquisition*. IRAL 41, 87-106.

while ours is in complex scripts (even if morphology is extremely simple, as in the case of Thai). We are in close contact; a letter of support and invitation to future collaboration from LDC Director Christopher Cieri is attached in Appendix 11.

4. LIKELIHOOD OF ACHIEVING RESULTS

The SEAlang Lab is the result of much planning and preparation. We began with extensive pre-development to ensure that our objectives were well-defined and realistic. Then, we widely demonstrated this advance work in open forums (including the ILR, CAST, and CASL) to invite critical commentary by language and software professionals. We believe that our methods are practical, and that our knowledge of the relevant issues in linguistics, computational linguistics, second-language acquisition, and software development are sufficient to enable success.

4a) Objectives clearly defined The SEAlang Lab will build tools that address a longstanding bottleneck in language instruction: the difficulty of reading, writing, and vocabulary acquisition typically encountered when learning complex-script, resource-scarce, LCTLs.

The Reader's, Writer's, and Vocabulary Workbenches are detailed elsewhere; functionality is described in Section 1a, and final format in Section 11. Examples can be seen in Appendix 1.

We will develop core vocabulary coverage lists and difficulty metrics; these are necessary elements of the Reader's and Vocabulary Workbenches, and are important in their own right for standards-based evaluation. We will coordinate this effort with the ongoing work of the Consortium for Advanced Study of Thai (CAST).

We will prepare additional lexical data as needed to support the Workbenches. This includes making SEAlang Library Thai-English dictionary coverage broader (adding new items promulgated by the Thai Royal Institute), deeper (improving existing coverage of usage, semantic relations, and word origins and ties), and more explicit (by listing derived noun or adjective forms). We are not rewriting the dictionary; but we are improving its fitness-to-task for our applications.

Our choice of Thai as an implementation language meets clear objectives. Thai manifests all the difficulties of complex-script languages, and thus lets us demonstrate the broadest possible

range of functionality, including automated segmentation. Its existing body of lexical resources, including dictionaries and mono- and bilingual text corpora, are sufficient to enable basic SEAlang Lab functions. Finally, it is a keystone language in a strategically important region.

Our Arabic-based Jawi script pilot implementation also meets clear objectives. Jawi is typical of Arabic scripts used to write non-Arab languages, including various Berber, Iranian, and Indo-Aryan languages. An excellent Pattani Malay / Thai / Malay dictionary is available (Dr. Christopher Court, 1984); it uses Thai and Rumi scripts and has been extended to Jawi.

4b) Methods practicable We have done considerable preliminary and proof-of-concept development in preparation for the SEAlang Lab (see also Appendix 1), seeking in particular:

- to assess the technical complexity and difficulty of the tasks involved,
- to determine how broad a range of functionality we can reasonably deliver,
- to begin a first design-implement-improve cycle,
- to demonstrate the project's potential to our colleagues in the LCTL community, and to seek both critical commentary and collaboration.

Key features of the SEAlang Lab's core software systems and critical data components have already been prototyped and demonstrated in basic form. While much work remains, we have shown that the SEAlang Lab's plan and implementation are practicable and achievable.

4c) Knowledge sufficient CRCL Inc. has already developed a broad range of lexical and technical resources relevant to Southeast Asian languages. In preparing for the this project, we have gone well beyond investigating the materials discussed in Section 3. We have:

- extensively reviewed the relevant literature on second language acquisition, with particular attention to research by Nation, Laufer, and Hill, and others in vocabulary acquisition,
- conducted broad surveys of related software development, including the work of Cobb, Kilgarriff, the REAP and LDC projects, and other similar efforts,
- consulted with a broad range of Southeast Asian and other language specialists, and become involved with CAST's ongoing evaluation-standards effort,
- assembled a team of specialists, including Doug Cooper (CRCL Executive Director and

SEAlang Library Technical Director), Prof. Robert Bickner (SEAlang Library Project Director and SEASSI Curriculum Director), Frank Smith (SEASSI Heritage Program Director, and author of many SEASSI course materials), Pongsorn Saipetch (software consultant), and Drs. Indrambarya and Suthiwan (consulting linguists); see Section 7b.

5. EXPECTED CONTRIBUTION TO OTHER PROGRAMS

The SEAlang Lab's contribution to Southeast Asian language study is discussed in Section 3.

Here, we look at the Lab's larger contribution to language competence, learning, and research.

In brief, the Reader's, Writer's, and Vocabulary Workbenches address closely related concerns:

- mechanical issues of reading and writing complex script / orthography languages,
- stylistic and syntactic issues that inhibit both receptive and productive competence,
- issues of vocabulary acquisition, retention, and automated drill production,
- other issues, including use of graded word lists to set difficulty levels for texts and tests.

The Workbenches allow in-depth focus on reading, writing, and vocabulary in isolation, or let us link their functionality – on-the-spot vocabulary drill to support reading, or the use of writing drills to support vocabulary retention – in innovative new ways. They are language independent, and can use any data that conforms to published standards for data markup.

Language learning Reading and writing are critical to developing fluency and/or professional competence. But all too often, they are taught using workbooks and methods meant for native speakers, as though mastering the shapes of letters were the only salient issue for literacy.

Unfortunately, as was discussed in Section 1a, complex scripts can make developing automaticity in reading far more difficult than simply learning to recognize letterforms. We do not really know very much about the best way to go about enabling automaticity, and indeed David Red (whose work we cited in Section 3a) has raised the prospect that it may be impossible for adult learners to *ever* become truly fluent readers of very different scripts.

Until we gain a better understanding of the pedagogy of complex scripts, it makes sense to explore the potential of assistive software tools in fostering automaticity. Thai script presents

many problems common to all complex scripts, so we think that lessons learned in Thai language studies will, in fact, be broadly useful. This experience can be applied to other languages by providing appropriate dictionaries and corpora for the Reader's and Writer's Workbenches.

Tools that dynamically generate tests and drills, in turn, will benefit any resource-poor LCTL. Seeing a definition once may be enough for native-language vocabulary building, but when studying a second language, it is not sufficient. As Nation and Laufer (cited in Section 3a) and others have demonstrated, cognitive effort must be expended in order to make new words stick.

The Vocabulary Workbench will stimulate that effort, whether it is used standalone or as part of a Reader's or Writer's Workbench session. Like the other tools, the Vocabulary Workbench does not promote any single approach to vocabulary acquisition. Rather, it provides as wide a range of exercises as we can generate automatically based on available lexical resources, along with the record-keeping facilities needed to measure performance.

Again, we think that lessons learned in the Thai context are not language-specific. There has been very little work done on dynamic, data-driven drill, and we expect that our experience in designing question types, extending lexical resources to support automatic drill generation, evaluating question difficulty by looking at the lexical content of alternative answers, and varying the timing and intensity of drill sessions will be broadly applicable.

Language competence Our tools will help translators, analysts, linguists, and other **language professionals** make the transition from professional competence to fluency, and help **heritage speakers** move from verbal fluency to adult-level literacy. They will also support new, flexible language-rating requirements. For example, the State Department is allowing alternatives to the standard 3-Reading / 3-Speaking / 3-Listening competence as part of its Language Learning Continuum initiative, letting a 2-Reading be offset by a 4-Speaking / 4-Listening rating. Assistive technology tools can help boost actual performance, despite the lower test rating.

The SEAlang Lab tools will also assist **academic** and **research users**, particularly in regions that have many mutually incomprehensible languages and scripts. For example, in Southeast Asia, movement in any direction, whether geographical or chronological, soon leads to new

language requirements. As Barbara Watson Andaya points out, “the range of linguistic and cultural groupings within Southeast Asia is so great that even those who have studied the region for an academic lifetime can only acquire real competence in a limited area.” (*Historian*, Spring 1995). The Reader’s and Writer’s Workbenches help scholars develop the regional language expertise that may be a prerequisite for area studies research.

Pedagogical research The SEAlang Lab confronts basic issues – reading, writing, and vocabulary – that are intensely researched in the context of English and the more commonly taught languages, but have practically been ignored in Middle Eastern and Asian language pedagogy.

For these languages, even the simplest pedagogical questions are rarely investigated. For example, in more than two decades of publication, the Federally funded journal *Reading in a Foreign Language* <http://nflrc.hawaii.edu/rfl> has not published a single article on South, Southeast, or Central Asian or Middle Eastern scripts. This is hardly a surprise; language teaching faculty in these areas are rarely hired for their research skills, or given the opportunity to develop them. We want to help create that opportunity.

The SEAlang Lab solves practical student problems, but it also opens fascinating research possibilities. This is no accident; most project participants have extensive teaching and research experience, and strong backgrounds in SEA linguistics. Thus, we provide tools that support alternative models (such as incidental *and* intentional vocabulary acquisition) and approaches (use of transcription *and* native scripts) with equal fervor. Partly it is because we know that different students, at different times, may benefit from all. More importantly, we want to **enable research** into questions that include:

- to develop true automaticity in reading, which is better: an early introduction to the local script, or an extended introduction using transcription or transliteration?
- can core wordlists be readily defined for the isolating languages typical of Southeast Asia? Are compound heads equivalent to lemmas when it comes to counting words?
- are reading skills transferable between mutually incomprehensible writing systems (like Burmese, Khmer, and Thai) that have the same underlying alpha-syllabic structure?

- can we automatically generate test and drill material that passes the Turing test, i.e. material that cannot be distinguished from a human teacher's work? Does it matter?
- what *is* the verdict on incidental vs. intentional vocabulary acquisition in complex-script languages when each is given the richest possible environment?

6. PLAN OF OPERATION

Project staff are listed below, and discussed in detail in Sections 6b (*management plan*) and 7a/b (*personnel*). The plan of operation is also discussed under Section 9b (*time schedule*).

<i>Position</i>	<i>Staff member</i>
Project Director / Software Developer	Doug Cooper
Lexical Resources Manager	Frank Smith
Data Developer	Namfon Buntua
Senior Curriculum Consultant	Professor Robert Bickner
Consulting Linguists	Professors Kitima Indrambarya and Titima Suthiwan
Software Consultant	Dr. Pongsorn Saipetch
Office Staff	Noosai Inthimas, Lwin Moe

6a) Project design The project's basic design is clear and manageable:

- Cooper will manage the project. He will also program and maintain the Workbenches, regularly adding content and functionality as lexical resources are extended by Buntua, and as core coverage lists, difficulty measures, and new drill types are developed by Smith. He will call on Saipetch for assistance on specialized software issues as needed.
- Smith will establish core coverage wordlists, develop text difficulty measures, and design new drill/exercise types. He and Cooper will work with Indrambarya and/or Suthiwan to specify lexical resource extensions required to support other functionality.
- Buntua will extend lexical resources, and assist Smith as necessary. Inthimas, also a native Thai speaker, will assist Buntua as required.
- Bickner is in charge of evaluation. He will advise and consult on tool development, manage relations with collaborating schools, and ensure fulfillment of our reporting and review requirements. Cooper and/or Smith will attend all collaborator workshops.

Note that the SEAlang Library (data) and SEAlang Lab (tools) are both integral components

of CRCL's long-term plan for SEA lexical resource development. By design, the SEAlang Library's outputs – mainly dictionary and corpus data – are plug-compatible with the SEAlang Lab's inputs (via a Web-accessible *application program interface*). This approach has allowed *rapid prototyping* of the SEAlang Lab systems we have already demonstrated, and paves the way for a smooth development path in the future.

6b) Management plan The project is small and easily managed by Cooper, who will work with Smith and Buntua in CRCL's Bangkok office, with linguist consultants Indrambarya or Suthiwan and software consultant Saipetch on call as necessary. Note that Smith resides in Bangkok for the full academic year.

As discussed in Section 9, *Evaluation*, Senior Curriculum Consultant Bickner (working in the U.S.) plays a special role in facilitating project oversight and evaluation. He will devote roughly one week per quarter to the project, focusing on the pedagogical content of our tools, and ensuring that we are addressing the needs and expectations of our collaborators.

6c) Relevance to 605 The SEAlang Lab directly addresses paragraph 6: *'the development and publication of specialized materials for use in foreign language, area studies, and other international fields, or for training foreign language, area, and other international specialists.'*

Section 605 also authorizes activities that *'contribute to achieving the purpose of this part.'* The SEAlang Lab tools are directly relevant to paragraph 5 (*'research on more effective methods of providing instruction and achieving competency in foreign languages'*), paragraph 7 (*'the uses of technology in foreign language, area studies, and international studies programs'*), and paragraph 8 (*'effective practices in the dissemination of international information, materials, research ... throughout the education community'*).

6d) Use of resources and personnel All staff have the skills required for their duties. Management of project staff is sufficient and appropriate, as is the support provided by office staff, and additional consulting for specialized linguist and computing assistance. See also Section 8a, **Budget Adequacy.**

6e) Equal access: The SEAlang Lab will be a universally accessible Web-based resource. We

will make every effort to comply with the **Web Accessibility Initiative**, which addresses the needs of those with disabilities (see also Appendix 6 and our GEPA 427 Compliance statement). Moreover, the Workbenches themselves are intended to make both language learning materials and foreign language texts as accessible as possible to all learners, and have particular relevance to American heritage learners of Southeast Asian ethnic extraction.

7. QUALITY OF PERSONNEL

The SEAlang Lab is a small, tightly focused project; it will be built in a research center dedicated exclusively to creating lexical resources and software tools for Southeast Asian languages. The applicant institution and project personnel are described here, while project collaborators are discussed in Section 10.

Applicant institution After nearly a decade of informal work in Bangkok as the *Center for Research in Computational Linguistics*, CRCL Inc. was registered in California as a 501(c)3 nonprofit in 2003. CRCL's mission is to promote research and study of the modern and historical languages of Thailand, Burma, Laos, Cambodia, and Vietnam, and to freely disseminate results to the international community of students and scholars (see also Appendix 2).

CRCL Inc. builds software systems that make Southeast Asian language data accessible for education and research, drawing on expertise in computer science, and extensive experience in SEA lexicography and linguistics. We currently focus on three projects: the SEAlang Lab, the SEAlang Library (which we initiated, and co-sponsor with U Wisconsin-Madison), and the Mon-Khmer Languages Comparative Dictionary project (which is managed by CRCL's Director of Mon-Khmer projects, Dr. Paul Sidwell).

7a) Project Director / Software Developer **Doug Cooper** is Executive Director of CRCL Inc., and Technical Director (0.4 FTE) of the SEAlang Library. He has written nine college-level texts and references, including the widely adopted *Oh! Pascal!*, and taught for five years in the Computer Science departments of UC Berkeley and Smith College. For the past decade he has focused on SEA computational linguistics research (including segmentation, optical character

recognition, and phonemic and orthographic approximation), lexicography, and Thai etymology, as well as digitizing and building tools for manipulating modern and historical texts. He manages the international listserv SEALANG-L, founded its lexicography group SEALEX, and has consulted for Microsoft and Monotype on Thai font design. He is Associate Editor of the forthcoming *H.L. Shorto Comparative Dictionary of the Mon-Khmer Languages* (Paul Sidwell, ed.).

7b) Key personnel **Frank Smith** is the SEAlang Lab's *Lexical Resources Manager* (0.5 FTE). He has a year-round 0.25 FTE appointment with SEASSI, where he is **Heritage Program Director** and **Khmer Instruction Coordinator**. Fluent in Khmer, Thai and Lao, he developed the course materials used by SEASSI's Khmer and Lao programs, trains Cambodian teachers to teach Khmer to foreigners, and helps direct SEASSI's teacher training program. A Bangkok resident (except during the 10-week SEASSI program), Mr. Smith's writing and teaching experience make him exceptionally qualified for his work in the SEAlang Lab, and his strong commitment to the region will help ensure that our work for Thai is carried on to other SEA languages.

Robert J. Bickner, the SEAlang Lab's *Senior Curriculum Consultant*, is **Professor of Thai Language and Literature** at UW-Madison. He is the **SEASSI Language Director**, and is Project Director of the SEAlang Library. One of the world's leading authorities on Thai linguistics and literature, Prof. Bickner supervises Thai language instruction at all levels, teaches courses in Thai linguistics, prose and poetry, and Southeast Asian literature in translation, and is Director of the *College Year in Thailand* program. He has published on management of, and performance measurement in, SEA language programs, and developed Thai-language computer-assisted learning (CALL) software. As Language Director of SEASSI, he is responsible for all program staffing and curriculum development for regular and heritage program courses.

Our **Consulting Linguists** both received their PhD's from the University of Hawai'i. Prof. Kitima Indrambarya (Kasetsart University) specializes in lexicase analyses of Thai verbs. Prof. Titima Suthiwan (National University of Singapore), our Pattani Malay specialist, wrote her thesis on Malay loanwords in Thai. Prof. Suthiwan spends a considerable amount of time with her family in Thailand throughout the year; scheduling consultation will not be a problem.

Our native-speaker **Developer** is Namfon Buntua, who did almost all of the development of the SEALang Library's Thai-English dictionary (based on the *Haas Thai Dictionary Project* data, which CRCL archives), and has worked with us on many other Thai-language projects.

Our **Software Consultant** is Dr. Pongsorn Saipetch. A native and resident of Bangkok, Dr. Saipetch earned his PhD from UCLA, and is founder of *Atrium Software*, a research-driven Bangkok-based software firm. Dr. Saipetch frequently consults for Thai government agencies, including the Council of State, the Office of the Criminal Court, and the Constitution Drafting Committee; his expertise in developing secure, high-performance Web services (as well as his experience in Thai-language software issues) make him particularly qualified for this project.

7c) Time commitment The project members' time commitments are adequate and appropriate. Cooper will have a 40% appointment on the SEALang Lab (he has an equal appointment on the SEALang Library project). Smith will devote 50% of his academic-year time to his SEALang Lab duties (he also has a 25% appointment with SEASSI). Bickner will devote approximately one week per quarter to the project. One native-speaker developer will work full time, with occasional assistance from the office manager if needed. Consulting linguists will be available for two weeks each quarter, and consultation on software for 15 days per year (see also Section 8a).

7d) Non-discrimination CRCL Inc. is committed to providing equal employment opportunities. Our linguist and software consultants, data developer, and office manager are Thai (including four women), and our computer systems manager is Burmese. We also subcontract considerable work (in digitizing dictionaries) to Digital Divide Data, a Cambodian / Lao firm created to provide employment for the disabled.

8. BUDGET AND COST EFFECTIVENESS

8a) Budget adequacy The budget requested for the SEALang Lab project is based on CRCL's experience in similar work, and on investigation (through extensive pre-development of SEALang Lab demonstration tools and data) of this project's needs. The budget is allocated primarily to working staff, as outlined below. Additional details are found in the Budget Narrative.

<i>Position</i>	<i>Support requested</i>
Project Director / Software Developer	40%
Lexical Resources Manager	50%
Data Developer	100%
Office Manager	40%
IT Systems Manager	40%
Linguist consulting	320 hours / year
Senior Curriculum Consultant	160 hours / year
Software Consultant	15 days / year

8b) Costs reasonable The SEALang Lab benefits greatly from the resources of the USED-funded SEALang Library, from the Bangkok field office and XML-tagged dictionary data that CRCL Inc. is funding as part of its Federal cost-share obligation to that project, and from the Thai audio files we have obtained for this project (see Appendix 9).

The fact that CRCL does not charge indirect costs also extends the project budget considerably. While the Library project was submitted by the University of Wisconsin-Madison (as required by TICFIA program regulations), we and our colleagues at Wisconsin jointly decided to submit the SEALang Lab proposal as CRCL Inc., a more cost-effective option that keeps our budget well within the International Research and Studies program guidelines.

9. EVALUATION PLAN

This year the Department of Education indicated the exceptional importance it places on accountability by tripling the number of points allocated to this section. We are responding in part by hiring a *Senior Curriculum Consultant* who has extensive experience in language teaching and project management, and who will take responsibility for ongoing project evaluation. Prof Robert Bickner fills this role. He will:

- ensure that project evaluation, assessment, and reporting are timely and constructive,
- ensure that the project's technical development is firmly grounded in sound pedagogy,
- ensure regular liaison with collaborators who are class-testing the SEALang Lab tools,
- devise protocols to compare learning outcomes supported by the SEALang Lab tools, when completed, to those achieved using traditional teaching and learning methods (such studies may be attempted by our collaborators during this project, but would more realis-

tically be conducted under a separate, follow-up research grant).

By design, the Senior Curriculum Consultant is not responsible for production of project software; as noted above, we allocate 4 weeks / year to this position. In addition, we will contract for an outside formal review as detailed below.

9a) Periodic evaluation Evaluation serves several functions, including ensuring that we are *a)* making regular progress in producing the SEAlang Lab deliverables, *b)* soliciting and responding to regular feedback on our work, and *c)* delivering results that meet the needs of all interested parties. We will perform this evaluation through these scheduled reviews:

- **Progress review** A Review Board, established to assess the University of Wisconsin's yearly performance on the SEAlang Library, will be asked to perform the same role for the SEAlang Lab. This board includes the Directors of CSEAS and SEASSI, and UW-Madison Languages and Cultures of Asia Prof. Ellen Rafferty.
- **Quality review** UW-Madison Prof. Robert Bickner will head an ongoing evaluation process that will elicit commentary and discussion from our collaborating partners at FSI, DLI, SEASSI, and CSEAS. We will hold regular workshops with each of our collaborators, focusing on tool design and use (year 1), teacher evaluation of tools (year 2), and student/teacher evaluation of tools (year 3).
- **Outside review** University of Washington Prof. Thomas Gething will provide a formal outside review at the end of year two. Prof. Gething is one of the most respected scholars in SEA Studies. As Director of the SEASSI program and Director of the Title VI-funded Advanced Study of Thai program, and as past Director of the University of Washington's Center for Southeast Asian Studies, Prof. Gething has deep and direct knowledge of the issues involved in SEA language pedagogy. As past Deputy Director of the National Foreign Language Center, and as a grant reviewer for the U.S. Department of Education and others, he has a broad understanding of American needs for language education.

Prof. Gething will evaluate our progress after 24 months, allowing us adequate time to follow his recommendations, and to apply them in class-testing, before final delivery.

- **Student/user evaluation** We will prepare and collect evaluation forms for students in our four pilot studies at FSI, DLI, SEASSI, and CSEAS. In addition, we will create an on-line form for anonymous Web users. All commentary will be forwarded to Professors Bickner and Gething, and will be summarized in the project's annual report.

9b) Time schedule The SEAlong Lab production schedule, and time allocations of principle staff members follows the plan of Section 6a, *Project design*. The overall project calendar is:

	quarter/year	4/06	1/07	2/07	3/07	4/07	1/08	2/08	3/08	4/8	1/09	2/09	3/09
READER'S WORKBENCH	display/segmentation	X				X				X			
	reference	X				X				X			
	coverage/difficulty				X			X			X	X	
WRITER'S WORKBENCH	pred'tive completion		X				X				X		
	usage comparison		X				X				X		
	learner's corpus										X	X	X
VOCABULARY WORKBENCH	test generation			X	X			X	X			X	X
	recordkeeping			X				X				X	
	save-test mode				X			X	X			X	
DATA DEVELOPMENT	core wordlists	X	X	X					X				
	difficulty metrics		X			X			X		X	X	
	data enrichment	X	X	X	X	X	X	X	X	X	X	X	X
EVALUATION	collab's workshop			X				X				X	
	outside review									X			
	internal evaluation				X				X			X	X

- **Reader's Workbench** Basic functionality will be enabled in the first six months, in time for spring 2007 discussion and summer 2007 class-testing. As enrichment of dictionary resources proceeds, new reference and display functions will be added to the tool in 2008 and 2009. Functionality related to core coverage and difficulty estimation will be added as developed. Retrieval of texts of specific difficulty / vocabulary coverage is planned for year 3.
- **Writer's Workbench** Basic functionality / class use are scheduled for 2007, with regular improvements enabled by lexical resource enhancements in subsequent years. We expect this tool to encourage greater emphasis on written course assignments, and will seek student / teacher permission to collect a *learner's corpus* of student assignments for error analysis.
- **Vocabulary Workbench** We will enable basic functionality for spring 2007 discussion and

summer 2007 student use. Features will be extended in subsequent years as they are suggested by classroom experience, and enabled by enriched lexicographic data. In years 2 and 3 we will use ‘teacher mode’ (create and save) in-house to prepare material for teacher discussion and evaluation via COTSEAL and the SEALteach listserv.

- **Core Vocabulary / Difficulty Metrics** Year 1 will develop rough coverage standards derived from corpus-based frequencies, basic text analysis, and discussion with teaching faculty. In year 2 these will be promulgated for discussion and evaluation, and integrated into the Reader’s Workbench. In year 3, we will test the proposal against texts that have been independently ranked for difficulty, and fine-tune coverage and difficulty metrics. This work will be coordinated with the *Consortium on Advanced Study of Thai* (CAST) yearly meeting.
- **Data Enrichment** Extension of lexicographic resources to support Workbench functionality will continue throughout the project, under the guidance of the Lexical Resources Manager during the academic year, and of the Project Director in the summer. A Linguist Consultant will periodically assist in specifying requirements and quality assurance on results.
- **Evaluation** Within the project, this is the primary responsibility of Prof. Robert Bickner. He will devote approximately one week per quarter to *a)* consulting with our collaborating partners, *b)* assessing faculty and student comments, *c)* collecting and reporting these to the Review Board / outside evaluator, along with his own ongoing assessment of project progress.

10. ADEQUACY OF RESOURCES

10a/b) Facilities and Equipment CRCL Inc. was created to pursue research and development projects involving Southeast Asian languages. We maintain a fully functional field office with broadband Internet access in Bangkok, Thailand, and contract for Web hosting in the U.S. Our computing facilities in Bangkok are excellent, as are our linguistics reference library resources. A full-time office manager and computer support engineer keep the facilities operating smoothly.

Additional expertise is readily available when needed. Our native-speaker consultants for Thai and Pattani Malay, Drs. Indrambarya and Suthiwan are highly regarded in the Thai aca-

demic community, as is software consultant Dr. Saipetch. Our CRCL Associates (see Appendix 2) can provide other specialized assistance if necessary. These include Dr. Stephen Morey (Tai languages), Dr. Paul Sidwell (Mon-Khmer languages), and James Clark (Director of the Open Source Software Division of the Thai government's Software Industry Promotion Agency).

11. DESCRIPTION OF FINAL FORMAT

Free, Web-accessible user interfaces will be provided for the Reader's, Writer's, and Vocabulary Workbenches. All development will conform to open, international standards for Web browsers, XHTML, XML, CSS, and JavaScript, as appropriate, and be operating-system independent. See Appendix 1 for examples of all tools described below.

Reader's Workbench The user will provide either a local file location, or an Internet URL. The page will be (invisibly) preprocessed and loaded. The user can apply three basic option sets: *a) presentation*, which allows for alternative fonts, segmentation into words or compounds, highlighting of particular kinds of words, etc., *b) reference*, which lets the mouse cursor and/or click present definitions, corpus or bitext examples, collocates, difficulty, and so on, and *c) evaluation*, which analyzes text difficulty, generates wordlists, and the like.

Linkage to the Vocabulary Workbench will provide either intensive (immediate follow-up drill on queried words) or extensive (delayed or no drill) reading. An embedded utility will let the user specify either a list of desired words, or a desired difficulty level, then retrieve locations of texts that match these requirements.

Writer's Workbench The user will have the option of entering *a) native orthography*, *b) phonetic transcription*, or *c) an English word found as part of a definition*. As the user types, a drop-down list will show candidate words, ordered either by length or by frequency in different text corpora. Words can be added to a running text, or selected to be checked against dictionaries, corpus or bitext examples, or known collocations. Multi-way contextual comparison (e.g. 'calamity' vs. 'disaster') will be provided.

An embedded utility will allow predictive completion at the phrase (rather than word) level,

providing phrase-completion capability. Once text has been entered, the user will be able to save it, or mail it, from within the Workbench.

Vocabulary Workbench The student user will specify a word, word list (such as core coverage, difficulty level, or the current week's lesson), or other source (such as words queried while using the Reader's Workbench). At least 25 exercise types will be provided, based on a variety of approaches to vocabulary drill (see examples in Appendix 1c). As exercises are generated and answered, the server will keep track of the student (anonymously), the words tested, and the exercise type and results. Students will have the option of specifying or disabling on-line record-keeping. If on-line records are kept, follow-up testing will be available. An embedded utility will let an instructor check, modify, and save automatically generated drills for use in class.

Other SEAlang Lab results Extensions and enrichments to the *Thai Dictionary Project* data will be accessible via the SEAlang Library tools, as part of the ordinary Thai-English dictionary resource. Our work on core coverage word lists, and our methodology for assessing text difficulty, will be published independently.

12. PROVISIONS FOR PRETESTING AND REVISION

Outside Review As discussed in Section 9, University of Washington Prof. Thomas Gething will conduct a formal outside review at the end of year 2, giving us adequate time to make improvements based on his recommendations in time for final submission to USED.

Class Testing SEAlang Lab tools will be pretested in four classroom environments,:

- **full-year intensive** programs of the DLI and FSI,
- **full-year non-intensive** programs of CSEAS,
- **summer intensive** programs of SEASSI,
- **summer heritage** programs of SEASSI.

As noted in Section 9, we will hold annual workshops at each institution to introduce the tools, discuss design and content issues, and learn from the experience of end-users. The software / data resources will be revised as necessary based on feedback provided by these meetings,

and on the comments of our reviewers. We will go to each institution, rather than fund national workshops, to reduce costs and ensure maximum participation.

The SEALang Lab tools (like the SEALang Library data resources) will be made available online for regular class and independent study use. Discussion with SEA language teaching faculty gives us every reason to believe that these tools will be adopted eagerly. The yearly COTSEAL conference and online SEALteach listserv will provide channels for additional feedback.

Collaborating Programs The **Foreign Service Institute** (FSI) is the Federal Government's primary training institution for officers and support personnel of the U.S. foreign service community, preparing American diplomats and other professionals to advance U.S. interests overseas and in Washington. At the George P. Shultz National Foreign Affairs Training Center, the FSI provides more than 450 courses, including some 70 that teach foreign languages, to more than 50,000 enrollees per year from the State Department and more than 40 other government agencies and the military service branches.

The **Defense Language Institute Foreign Language Center** (DLIFLC) is the world's largest language institute, and is regarded as one of the world's finest schools for foreign language instruction. It educates, evaluates, and supports foreign language specialists and provides the Department of Defense and other Federal agencies with linguists able to support United States interests worldwide. Resident instruction is provided at the Presidio of Monterey in some 23 languages plus several dialects, and can accommodate approximately 3,500 students.

The **Center for Southeast Asian Studies** (CSEAS) at the University of Wisconsin-Madison is one of eight National Resource Centers for the study of Southeast Asia. Known worldwide for innovation and excellence in language instruction, UW-Madison is also home to the **National Council of Organizations of Less Commonly Taught Languages**, as well as a **Language Institute** that offers a doctoral program in second language acquisition.

With more than 25 years of commitment to SEA language and area studies, CSEAS hosts the internationally recognized **Southeast Asian Studies Summer Institute**. Established in 1983 by 14 North American partners, SEASSI is the only national organization of American SEA Studies

centers; it provides intensive language instruction in at least eight major SEA languages each year. A leading international expert in each language plans and supervises the curriculum; teaching staff are all native speakers and highly experienced second-language teaching professionals.

13. INVITATIONAL PRIORITIES

Priority 1 *The development of instructional materials for use by students, teachers, and college faculty that focus on Islamic societies and the languages of those societies.*

Not all societies are defined by national borders. Such is the case in Thailand, whose Southernmost provinces – formerly independent sultanates annexed by Thailand in 1902 – have overwhelmingly Muslim populations, and an enduring Malay / Islamic self-identity. The **Pattani Malay** dialect is spoken throughout the region, and its **Arabic-based Jawi script** is universally taught in the more than 500 increasingly popular Pondok schools (see also Appendix 5).

Sadly, a near-daily cycle of bombings, beheadings, and draconian government response have claimed the lives of well over 1,000 people since January, 2004. Martial law was recently replaced with rule by decree, which suspends many civil rights, and grants immunity from prosecution to state security forces. The violence shows no signs of diminishing, leaving Thailand locked in an intense battle to determine whether, and how, it will come to terms with the fact that its Southernmost provinces form an internal – and increasingly violent – Islamic society.

The SEAlang Lab tools will manage both the **Arabic-based Jawi** and non-segmented Thai scripts used to write Pattani Malay (which is not currently taught in the U.S.). Resources will include a Pattani Malay / Thai / English dictionary, and facilities that allow conversion between Jawi, transcribed Thai script, and the roman-alphabet Rumi script used in Malaysia.

The SEAlang Lab Reader's Workbench will also enhance access to **Thai-language primary sources**, such as on-line newspapers and discussion groups, by American teachers and researchers. By enabling the use of authentic texts (in both Thai and Jawi) as instructional materials, the SEAlang Lab will make a clear contribution to America's understanding of conflict at the intersection of Southeast Asia's Buddhist and Islamic worlds.

TABLE OF CONTENTS

Narrative

1. NEED FOR THE PROJECT	1
2. USEFULNESS OF THE PRODUCT TO OTHERS.....	6
3. ACCOUNT OF RELATED MATERIALS	9
4. LIKELIHOOD OF ACHIEVING RESULTS.....	13
5. EXPECTED CONTRIBUTION TO OTHER PROGRAMS.....	15
6. PLAN OF OPERATION.....	18
7. QUALITY OF PERSONNEL	20
8. BUDGET AND COST EFFECTIVENESS	22
9. EVALUATION PLAN.....	23
10. ADEQUACY OF RESOURCES	26
11. DESCRIPTION OF FINAL FORMAT	27
12. PROVISIONS FOR PRETESTING AND REVISION	28
13. INVITATIONAL PRIORITIES.....	30

Appendices

Appendix 1. Examples of SEAlang Lab Preliminary Development	A1
<i>1a) Reader's Workbench</i>	A1
<i>1b) Writer's Workbench</i>	A6
<i>1c) Vocabulary Workbench</i>	A8
Appendix 2. CRCL Inc. and CRCL Associates	A13
Appendix 3. The SEAlang Library	A15
Appendix 4. Examples of SEAlang Library Preliminary Development	A18
Appendix 5. Jawi Script	A23
Appendix 6. Web Accessibility Initiative	A28
Appendix 7. Cooperating Institutions	A31
Appendix 8. Resumes of Key Personnel	A36
Appendix 9. License to Use Thai Audio Lexicon	A57
Appendix 10. Letters Confirming Collaboration	A58
Appendix 11. Letters of Support	A63

Project Narrative

Other Narrative

Attachment 1:

Title: Pages: Uploaded File: **6203-Mandatory_irs05-appendix.pdf**

Appendices

APPENDIX 1. EXAMPLES OF SEALANG LAB PRELIMINARY DEVELOPMENT.....	A1
1A) READER'S WORKBENCH.....	A1
1B) WRITER'S WORKBENCH.....	A6
1C) VOCABULARY WORKBENCH.....	A8
APPENDIX 2. CRCL INC. AND CRCL ASSOCIATES.....	A13
APPENDIX 3. THE SEALANG LIBRARY.....	A15
APPENDIX 4. EXAMPLES OF SEALANG LIBRARY PRELIMINARY DEVELOPMENT.....	A18
APPENDIX 5. JAWI SCRIPT.....	A23
APPENDIX 6. WEB ACCESSIBILITY INITIATIVE.....	A28
APPENDIX 7. COOPERATING INSTITUTIONS.....	A31
FOREIGN SERVICE INSTITUTE (FSI)	
DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER (DLI/FLC)	
UNIVERSITY OF WISCONSIN-MADISON CENTER FOR SOUTHEAST ASIAN STUDIES (CSEAS)	
Southeast Asian Summer Studies Institute (SEASSI)	
APPENDIX 8. RESUMES OF KEY PERSONNEL.....	A36
DOUG COOPER, Project Director and Software Developer	
PROF. ROBERT J. BICKNER, Senior Curriculum Consultant	
FRANK SMITH, Lexical Resources Manager	
PROF. THOMAS GETHING, Outside Reviewer	
KITIMA INDRAMBARAYA, Linguist Consultant	
TITIMA SUTHIWAN, Linguist Consultant	
PONGSKORN SAIPETCH, Software Consultant	
LWIN MOE, IT Systems Manager	
APPENDIX 9. LICENSE TO USE THAI AUDIO LEXICON.....	A57
APPENDIX 10. LETTERS CONFIRMING COLLABORATION.....	A58
DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER	
FOREIGN SERVICE INSTITUTE	
UNIVERSITY OF WISCONSIN-MADISON CENTER FOR SOUTHEAST ASIAN STUDIES / SEASSI	
APPENDIX 11. LETTERS OF SUPPORT.....	A63
COORDINATOR, INTERAGENCY LANGUAGE ROUNDTABLE	
PRESIDENT AND VICE-PRESIDENT, COUNCIL OF TEACHERS OF SOUTHEAST ASIAN LANGUAGE	
EXECUTIVE DIRECTOR, NATIONAL COUNCIL OF LESS COMMONLY TAUGHT LANGUAGES	
EXECUTIVE DIRECTOR, LINGUISTIC DATA CONSORTIUM	
DIRECTOR, SOUTHEAST ASIAN STUDIES SUMMER INSTITUTE	

Appendix 1. Examples of SEAlang Lab Preliminary Development

CRCL Inc. has done considerable preliminary design work and proof-of-concept development in preparation for the SEAlang Lab, with the intention of:

- assessing the technical complexity and difficulty of the tasks involved,
- determining how broad a range of functionality we might reasonably deliver,
- beginning a first design-implement-improve cycle,
- demonstrating the project’s potential to our colleagues in the LCTL community,
- seeking critical commentary on our plan of operation.

The screen captures, below, show the Reader’s, Writer’s, and Vocabulary Workbenches as displayed on an ordinary Web browser (Firefox, but Netscape, Safari, and other standards-compliant browsers work in exactly the same way). All these examples are taken from presentations made in 2005, and all tools actually work as shown. These demonstrations were intended to display a broad range of functionality to audiences of language and software professionals – these are *not* the final designs that will be used in the classroom or made available on-line.

1a) Reader’s Workbench

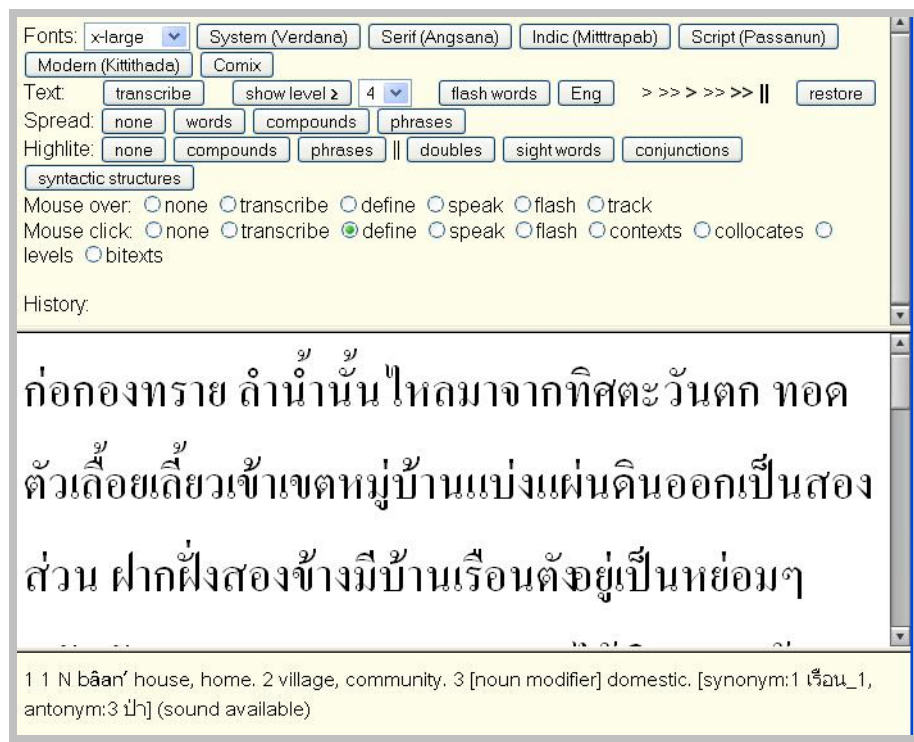


Figure 1. A rapid prototype of the Reader’s Workbench, after clicking on the word บ้าน ‘house,’ located roughly in the middle of the text sample (note that the definition appears at the bottom). On top, we integrate various kinds of functionality from the SEAlang Library (including dictionary and corpus lookup), plus many additional features designed just for Workbench applications. Note the circular ‘head’ that is found at the start of most letters, and which is universally taught as the key feature of standard Thai penmanship, and compare to letters in Figure 2, below.

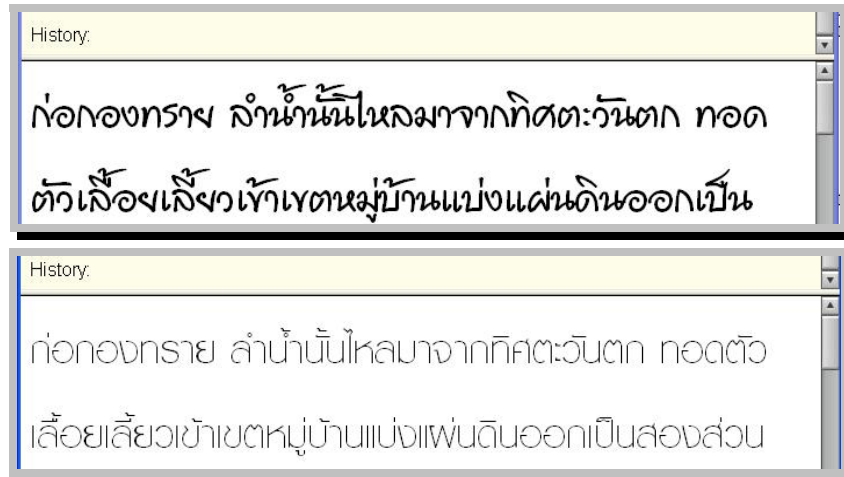


Figure 2. The importance of presentation (here, showing the text of Figure 1 in script and modern fonts). In the real world, Thai letters often don't have the circular heads seen in Figure 1. Nor is it a trivial matter to learn how to read new font styles. This becomes a serious issue when students use the authentic materials, from comic books (script fonts) to glossy magazines (modern fonts) that are becoming increasingly available. The Reader's Workbench allows transformation of texts that are available in electronic form, or training to handle texts that are not.

(The word **History** at the top of these screen captures is explained in Figure 7, below.)

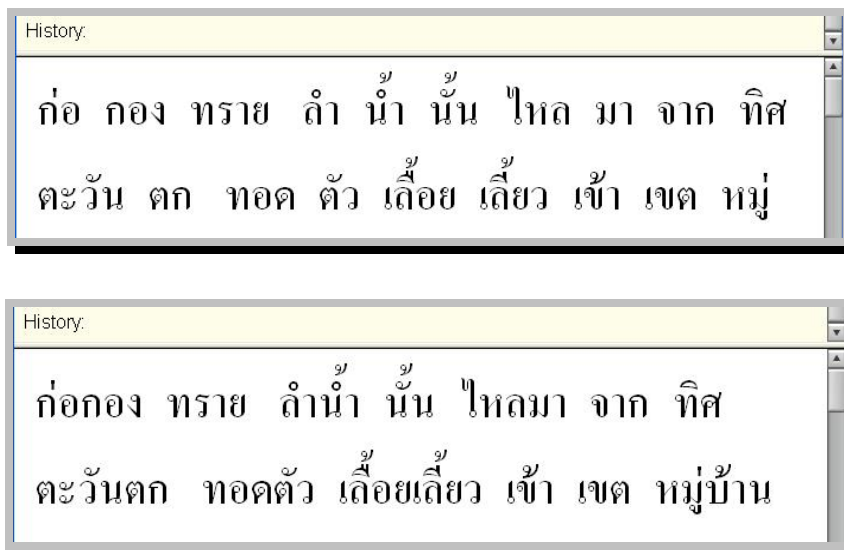


Figure 3. Thai text from Figure 1, segmented by the Reader's Workbench for easier reading. The upper image shows text broken into individual words – the result of standard text segmentation algorithms, and dictionary data supplied by the SEALang Library. The lower image shows SEALang Lab extensions: the text is broken into compounds, based on a combination of improved algorithms, and extended / enriched dictionary content.

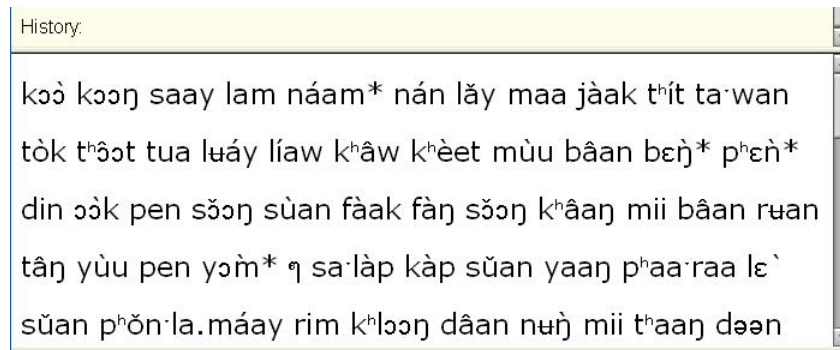


Figure 4. Here, the same text is segmented and then phonetically transcribed, based on data from the SEAlang Library's Thai-English dictionary. Words marked with an asterisk have irregular pronunciation. This tool is immediately practical for both individual and classroom use – in fact, a Workbench utility will let text be transcribed and saved like this automatically. At the same time, it provides the instrumentation we need to explore one of the most fundamental research questions of complex-script pedagogy: are students better off with early exposure to native orthography, or would they benefit more by building substantial vocabularies – which they might gain by covering larger amounts of phonetically transcribed text – first?

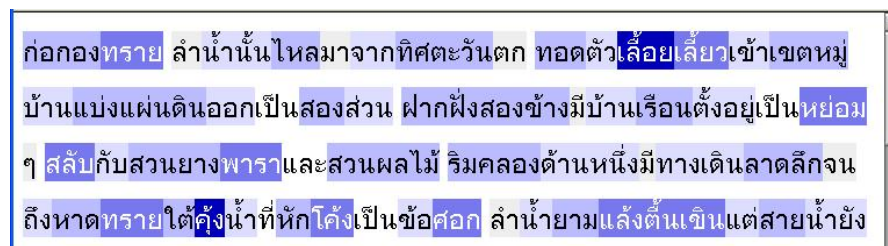


Figure 5. Here, words are shaded by difficulty level (using a very crude measure based on word frequency in a text corpus). In practice, highlighting just the least common words (the few words with the darkest shading) is enormously helpful to students who are reading nonsegmented complex-scripts, because they make it possible to read the unknown word's context. Without highlighting, it can be difficult even to figure out where the new word begins and ends. This is extremely tiring; a tiny bit of assistance makes most students more willing to read much more text.



Figure 6. These display options will make more sense now. The Reader's Workbench goes far beyond traditional glossing in marking up texts for student use. And, it serves readers at a far wider range of abilities – even a reader who is fairly adept at reading nonsegmented text may wish to have text broken into compounds or phrases when tackling an unusual new font (like those of Figure 2), or an exceptionally difficult or colloquial writing style.



Figure 7. The mouse and keyboard can be associated with various programmed (or Web-accessible) actions. What is more interesting, though, is that we can automatically keep track of mouse movement and clicks for later use, by asking the reader to use the mouse like a pointing finger. Here, “mouse over – track” means ‘track the amount of time the reader spends in the neighborhood of each word.’ Later, this lets us know if the student paused to read a word’s definition, or stumbled and had to backtrack while reading. We can use this information later when we automatically generate drill questions for the student. Recording each mouse click provides the same sort of information, and tells us more about how the Workbench is actually being used.

Above, the ‘0.5’ on the third line indicates that the cursor was near the word สอง for about half a second (a fairly long time). Only the most recent word/time are shown, but all are recorded. The **History** line shows which functions were called for which words.

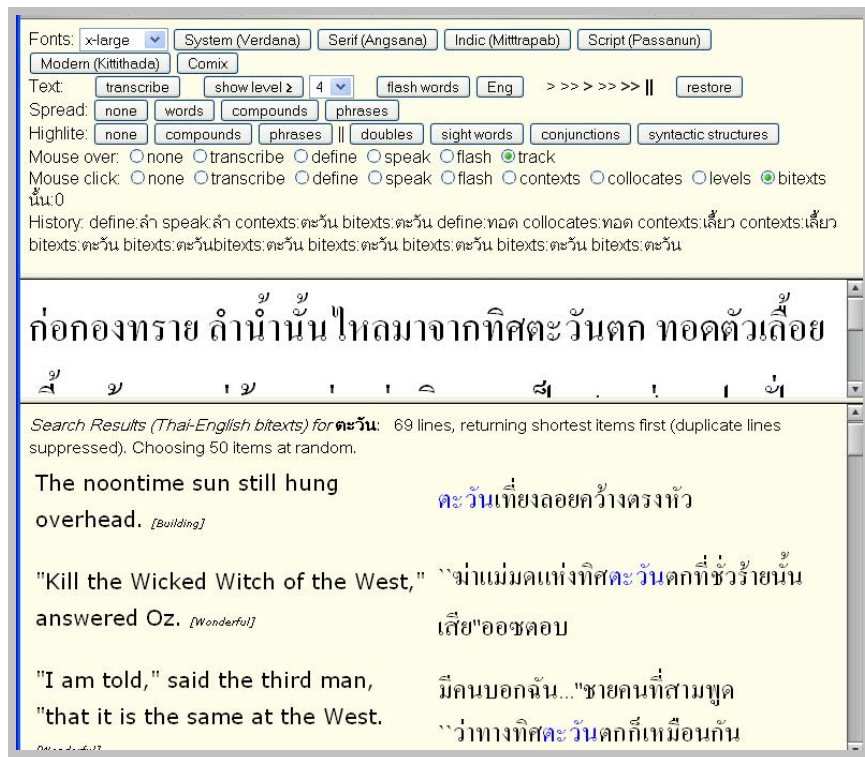


Figure 8. The Reader’s Workbench in action – the reader clicks on a word to see it in other translated contexts. The bitext data is from the SEALang Library bitext corpus, which can also be invoked as a separate standalone Web service. These same bitext contexts will also be used in other applications, such as generating drill and test.



Figure 9. The real power of the Reader's Workbench is that it can be used as a window to any and all Web content. Here is an inside page of the Thai daily newspaper *Thai Rath*. Note that links within pages continue to work properly.



Figure 10. The same page after clicking **transcribe**, which segments and phonetically transcribes the text (including links). Some words are still in Thai script because they were not found in the current dictionary. Improving the dictionary to allow proper transcription of Thai numbers, letters, and common abbreviations is one example of the sort of dictionary extension the SEALang Lab will provide.

1b) Writer's Workbench

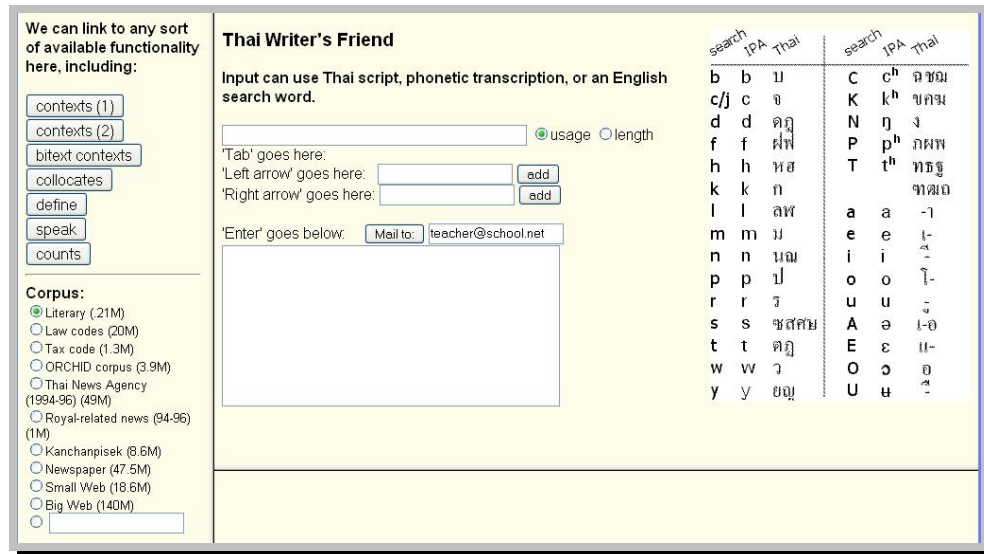


Figure 11. The Writer's Workbench (demonstrated here as the *Thai Writer's Friend*). This is also a prototype system meant to display potential functionality; it is not laid out in final user form. On the lower left, note that the tools can be associated with different text corpora, which can be significant for different styles and text coverage.

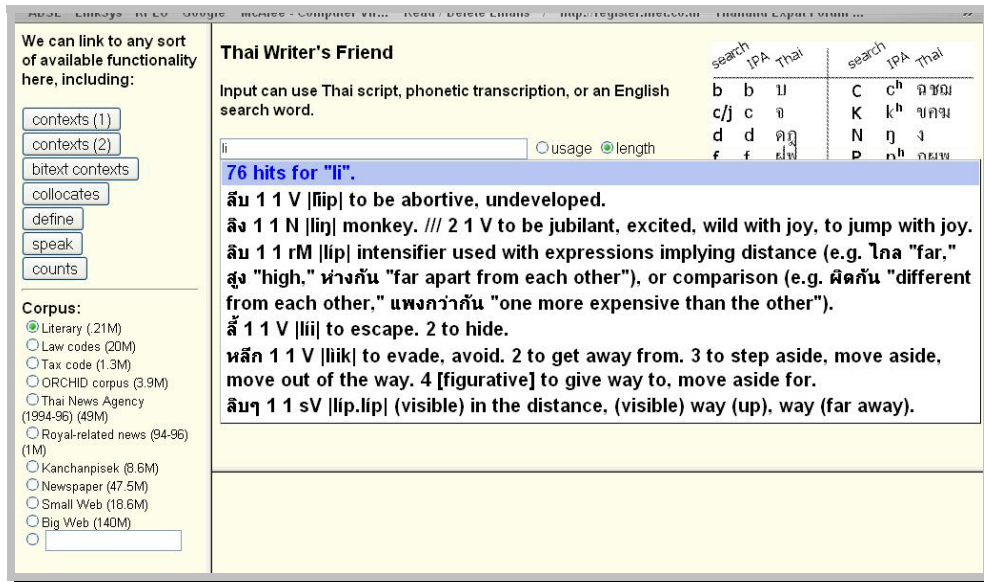


Figure 12. Here, the user searches for the Thai word for copyright, /likkhasit/, using romanized phonetics starting with *li*. The Workbench instantly fills the drop-down list after each letter, using the same underlying Web server technology as *Google Suggest*. Here, the Workbench uses length (shortest possibility first) as its predictive completion model. This example shows the full definition, including IPA transcription, but the drop-down can be limited to display just Thai text.

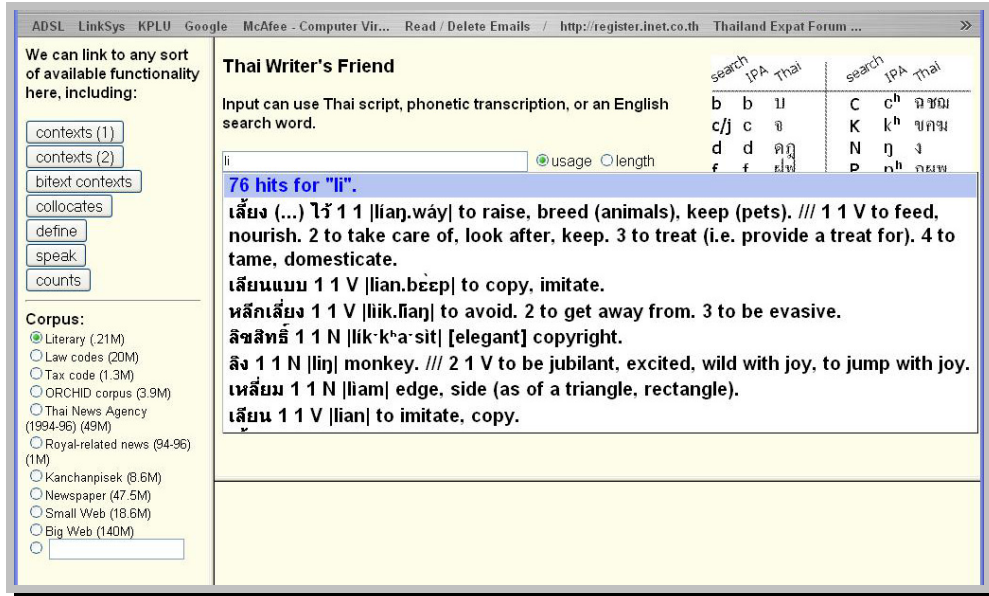


Figure 13. The same entry, but using frequency in an open corpus of Web text as the model for predictive completion. The word we seek (Thai ลิขสิทธ์) is the fourth most-frequent entry (because many Web pages have copyright notices). As it happens, this word has an extremely irregular Thai spelling that would be difficult even for advanced students to intuit correctly.

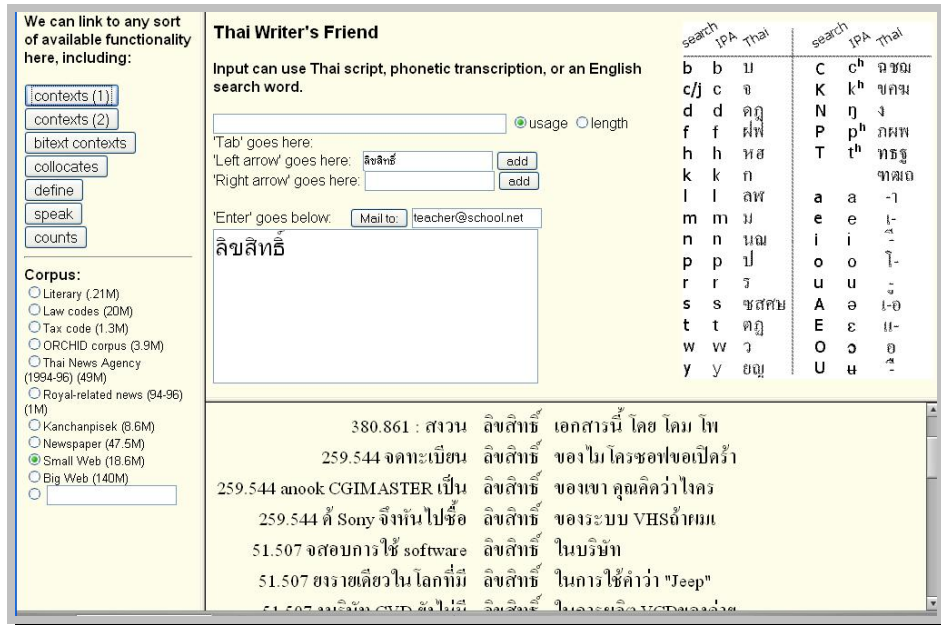


Figure 14. Once a word is selected, it can be added to a running text with the 'Enter' key, or saved and sent to one of the other SEALang Library or SEALang Lab tools. Here, we see examples of the item we selected in context. The Workbench can also complete words that are entered in Thai script.

we can link to any sort of available functionality here, including:

- contexts (1)
- contexts (2)
- bitext contexts
- collocates
- define
- speak
- counts

Corpus:

- Literary (21M)
- Law codes (20M)
- Tax code (1.3M)
- ORCHID corpus (3.9M)
- Thai News Agency (1994-96) (49M)
- Royal-related news (94-96) (1M)
- Kanchanpisek (8.6M)
- Newspaper (47.5M)
- Small Web (18.6M)
- Big Web (140M)

Thai Writer's Friend

Input can use Thai script, phonetic transcription, or an English search word.

copyright usage length

2 hits for "copyright".
ลิขสิทธิ์ 1 1 N |lik·k'ha·sit| [elegant] copyright.
กรรมสิทธิ์ 1 1 N |kam·ma·sit| ownership. 2 monopoly. 3 rights, copyright.

'Enter' goes below:

ลิขสิทธิ์

search	IPA	Thai	search	IPA	Thai
b	b	บ	C	c ^h	ช ชฌ
c/j	c	จ	K	k ^h	ข คฌ
d	d	ด	N	ŋ	ง
f	f	ฝ ฟ	P	p ^h	ภ ผ พ
			t ^h		ท ฐ
			a	-	า
			m	m	ม
			n	n	น ณ
			p	p	ป
			r	r	ร
			s	s	ซ ส ศ สษ
			t	t	ต ถ
			w	w	ว
			y	y	ย
			e	e	เ -
			i	i	อี -
			o	o	โ -
			u	u	อ -
			A	ə	เอ -
			E	ɛ	แ -
			O	ɔ	อ -
			U	u	อุ -

16 distinct leading collocates:
 () กรรมสิทธิ์ (1) มีกรรมสิทธิ์ (7) เป็นกรรมสิทธิ์
 ===== 25% (top 3 items = 18 of 59 hits)
 (6) คือกรรมสิทธิ์ (5) ได้กรรมสิทธิ์ (2) แสดงกรรมสิทธิ์
 ===== 50% (top 6 items = 31 of 59 hits)
 (2) ถึงกรรมสิทธิ์ (2) เจ้าของกรรมสิทธิ์ (1) ยกกรรมสิทธิ์
 ===== 60% (top 9 items = 36 of 59 hits)
 (1) เอกสารกรรมสิทธิ์ (1) ส่วนแบ่งกรรมสิทธิ์
 (1) เสียกรรมสิทธิ์ (1) --กรรมสิทธิ์ (1) แห่งกรรมสิทธิ์
 (1) เปลี่ยนกรรมสิทธิ์

25 distinct trailing collocates:
 () ในกรรมสิทธิ์ () ของกรรมสิทธิ์
 ===== 25% (top 2 items = 19 of 59 hits)
 () ที่ดินกรรมสิทธิ์ () อนุกรรมสิทธิ์ () บ้านกรรมสิทธิ์
 () ไร่กรรมสิทธิ์
 ===== 50% (top 6 items = 31 of 59 hits)
 () ยานพาหนะกรรมสิทธิ์ (1) แห่งกรรมสิทธิ์
 () ครั้งหนึ่งกรรมสิทธิ์
 ===== 60% (top 9 items = 37 of 59 hits)
 (5) ได้กรรมสิทธิ์ () ร่วมกับกรรมสิทธิ์ () ครั้นกรรมสิทธิ์

Find: Find Next Find Previous Highlight Match case

Figure 15. Above, a reverse lookup for ‘copyright.’ I’ve saved the second entry, /kammasit/, and show its immediate collocates at the bottom of the screen capture. The tool can compare the collocational features of two similar words (an English example might be *disaster* versus *calamity*) as well.

1c) Vocabulary Workbench

The Vocabulary Workbench automatically generates drill and test questions, and was previewed by demonstrating more than two dozen question types. In the interests of comprehensibility, many captures below show synthetic English examples; Thai equivalents are easily generated. The primary goal of the Vocabulary Workbench is to increase the cognitive ‘learning load’ associated with new vocabulary; not to competitively rank the student’s academic performance.

What's the missing collocate?

on- -truck -house -man

What's the missing collocate?

on truck house man

Figure 16. Collocates can be extracted by analyzing text corpora, or simply taken from the compound word entries that dominate Southeast Asian dictionaries. Then, we use this information to generate drill questions. Above, a simple variation – do we or don’t we mark the missing word’s position? – creates easy and difficult variations of the same problem. The missing word is **fire**.

Which four are likely collocates of "fire"?

green run **on** **truck** singer arithmetic **house** man

Figure 17. A variation on the same problem, produced by adding randomly chosen distractor terms, **man** hasn't been clicked yet. Making the question a little bit easier (*Which three ...*, rather than four), and/or looking at the difficulty level of the true collocates makes such problems more robust.

Re-order the sentence (window is +/- 1 word)

I rushed home, fire the but had already spread out of control.

Many cities have safety fire developed programs.

Re-order the sentence (window is +/- 1 word)

I rushed home, **but** fire the had already spread out of control.

Many cities have safety fire developed programs.

Figure 18. The drill word is **fire**. The Workbench finds corpus examples that contain this word, then reorders a small 'window' around it: *but the fire* becomes *fire the but*. In the second illustration, the user has clicked on **but**, which is automatically moved into the correct position.

Where does the missing word "fire" go?

but when I got there the **fire** had already spread beyond
announced plans to more than 3,000 other employees
site provides information on data, safety, prevention, public education

Figure 19. An automatically generated unmarked cloze test. The user has already clicked in the proper place to reveal the correct position for **fire**.

Which is the definition of "fire"?

- the number between 3 and 5
- equitable; an open-air festival
- a flaming or burning ...
- dread, fright; a feeling of ...;

Figure 20. Each choice – *four, fair, fear, fire* – defines a minimal phonemic variation. The SEAlang Library dictionary has the data we need to generate this kind of drill automatically.

Which is the translated sentence?

The **house** must have fallen on her.

- ฉันจะทำบ้าน ... เธอพูดกับ ตัวเอง
- บ้านคงตกลงมาทับเธอแน่ๆ
- เจ้าของบ้านตกใจร้องลั่น
- ดูสิ"เรอกสวาทอ แล้วชี้ ไปที่ มุมบ้าน

Which is the translated sentence?

บ้านคงตกลงมาทับเธอแน่ๆ

- The owner of the sand **house** cried out.
- I'll make a **house**, she told herself.
- See! she continued, pointing to the corner of the **house**.
- The **house** must have fallen on her.

Figure 21. Aligned pairs drawn from a bitext corpus are readily turned into drills. Once word-coverage data is available, we will be able to assess the difficulty of potential match sentences and provide L2 choices that are roughly equivalent.

Find each instance of
this word:

these words: **ทราย ผิง ต้นเขิน**

the next word you hear:

กอกองทราย สำน้ำนั้นไหลมาจากทิศตะวันตก ทอดตัวเลี้ยวเลี้ยวเข้าเขตหมู่บ้านแบ่งแผ่นดินออกเป็นสอง
ส่วน ผากผิงสองข้างมีบ้านเรือนตั้งอยู่เป็นหย่อมๆ สลับกับสวนยางพาราและสวนผลไม้ ริมคลองด้านหนึ่งมี
ทางเดินลาดลึกลงถึงหาดทรายใต้ค้ำน้ำที่หักโค้งเป็นข้อศอก สำน้ำยามแล้งต้นเขินแต่สายน้ำยังคงเอื่อยริน

Figure 22. The highlighted words pop up when the student clicks on *these words*, similarly, one or more words are spoken when *the next word you hear* is clicked.

Which is the proper complete verb?
 She drank the liquid

Figure 23. Thai and other Southeast Asian languages require verbs of *completion* or *directionality*. Although these are not usually given in dictionaries, it is a straightforward task to add them, and use them as the basis of automatically generated drill. These are typical of the ways we will extend existing dictionary data.

Which is the most appropriate verb for everyday use?
 She the liquid.

Figure 24. Social register is a particularly important distinction in Thai, only partially captured by comparison to English *imbibed*, *drank*, *guzzled*. Synonym sets based on social register are a very helpful source of drill question types; again, extending the dictionary appropriately is a straightforward matter.

Which word do you hear (tone contrast)?
 [1] [2] [3] [4]
 A. ฟน B. ฟล C. ฟน D. ฟล

Which word do you hear (leading consonant contrast)?
 [1] [2] [3]
 A. ฟน B. ฟน C. ฟน

Figure 25. The SEALang Lab has been given permission to use an extensive Thai audio lexicon, and sound files underlie many of the problem types. We can automatically choose sound files that test the proper minimal contrast because the SEALang Library includes phonetic transcription of every Thai word in its dictionary.

Click me, then retype this word: *perfect!*

Listen, then type this word: *so far so good*

Figure 26. The upper example flashes one or more words briefly, the lower example plays a sound file. The student can be given letter-by-letter feedback (with explicit detail; e.g. *No, try a vowel*) as necessary. The goal is not to teach typing, although the tool can be used for that. Rather, it is to build the student's visual and phonemic memory for very unfamiliar script and sound systems.

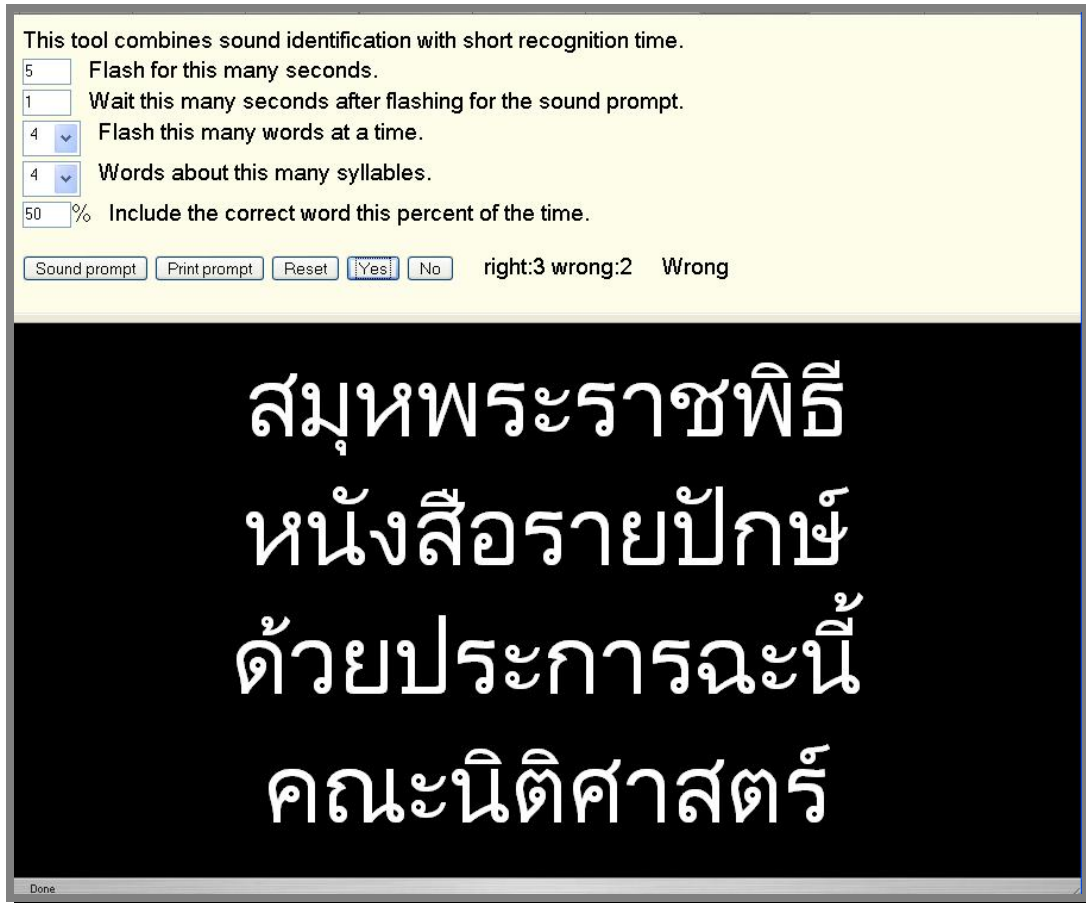


Figure 27. This tool ties sound to reading. Here, the student briefly sees four four-syllable words or phrases, then waits to hear a spoken phrase which may or may not have been included. The test can be reversed – sound first, then print – and all parameters can be varied for testing purposes, with times measured reliably below one tenth of a second. Audio samples can be chosen at random from the 24,000-entry audio lexicon we have obtained for use in the SEALang Lab; we analyze their syllable structure using data from the SEALang Library’s Thai-English dictionary.

This particular example, like the others, is simply intended to demonstrate the kind of functionality that the SEALang Lab can deliver. Nevertheless, it is clear that we can create not only practice and test tools, but also research instrumentation that will enable study of basic questions about the development of automaticity in reading complex-script languages.

APPENDIX 2. CRCL INC. AND CRCL ASSOCIATES

The Center for Research in Computational Linguistics was established informally in Bangkok in 1995, and formally registered as CRCL Inc., a US 501(c)(3) nonprofit organization, in 2003. Formed for educational purposes under the California Nonprofit Public Benefit Corporation Law, CRCL Inc. has been granted federal and state tax-exempt status.

CRCL's mission is to promote research and study of the modern and historical languages of Thailand, Burma, Laos, Cambodia, and Vietnam, and to freely disseminate findings and results to the international community of students and scholars. We pursue applications in:

- **education**, including development of tools, resources, and methodology for instruction in these less-commonly taught languages; and
- **linguistics research**, including research in etymology, computational linguistics, and computational lexicography, as well as related disciplines and research.

CRCL's core competence lies in writing software that makes language data accessible; our competitive advantage in research and development derives from our expertise in computer science, and extensive experience in SEA lexicography and linguistics. Besides pursuing our own research agenda, the CRCL Inc. outreach program provides technical assistance to scholars who, for lack of funding or expertise, cannot adequately develop or promulgate their results.

Associates: CRCL Inc. *Associates* are established professionals in linguistics and computer science who support our mission, share our interests, and from time to time consult for or advise us on topics of mutual concern. All of our associates spend a considerable amount of time each year in Southeast Asia, typically doing fieldwork in linguistics, or open-source software development. They include:

Mark Alves (Ph.D. U Hawai'i) Dr. Alves specializes in Vietnamese linguistics and second language acquisition, with additional interests in Chinese and the minority Mon-Khmer languages of Vietnam. He has published extensively on Vietnamese historical linguistics, focusing on the origins of the Vietnamese language, the history of Vietnamese-Chinese language contact, and the linguistic typology of Vietnam as a borderline Southeast Asian/Chinese language. He is currently exploring issues of grammaticalization within Vietnamese, with particular attention to Sino-Vietnamese grammaticalized vocabulary.

James Clark Founder of the *Thai Open Source Software Center*, he has been involved with SGML and XML for more than 10 years. He has contributed to many standards, creating a considerable body of open source software, and was awarded the industry's first *XML Cup*; was technical lead of the *XML 1.0 Recommendation*, editor of the *XPath* and *XSLT Recommendations*, and main author of the *DSSSL* (ISO 10179) standard. Currently, he is chair of the *OASIS RELAX NG TC* and editor of the *RELAX NG* specification. He also founded, and supports, a non-profit K-12 school intended to provide new models for rural Thai education.

Allan Cooper (Ph.D. MIT) Dr. Cooper researched mathematical models for many years at the Los Alamos National Laboratories before moving on to software modeling and design with Microsoft Inc. In 1993 he founded *Digital Word*, a software research and consulting company in Bellevue, Washington, that specializes in the design of Web-accessible data resources. Dr Cooper has been closely associated with CRCL since its inception.

Richard Kunst (Ph.D. UC Berkeley) Dr. Kunst earned his doctorate in Chinese Language and Literature. He taught both Chinese and Japanese language and literature at Duke University (1975 – 1991); he is also Executive Director of the Humanities Computing Laboratory, the successor to the DUCALL (Duke University Computer-Assisted Language Learning) Project, and is Adjunct Professor of Asian Languages at Duke University. At present Dr. Kunst is technical lead on the School of Oriental and African Studies *Wa Dictionary & Internet Database for Minority Languages of Burma* project.

Stephen Morey (Ph.D. Monash U) Dr. Morey is a Research Fellow of the *Research Center for Linguistic Typology*, Latrobe University. His research interests focus on the Tai languages, particularly the Shan-related languages found in the Assam region of Northeastern India, as well as Turung, long thought to be Tai-related, but actually a Tibeto-Burman language similar to Singpho. Dr. Morey has also done extensive work on font design, and on the initiation of printing and reading programs for endangered languages in the region.

Pongsorn Saipetch (Ph.D. UCLA) Dr. Saipetch is founder of *Atrium Software*, a research-driven Bangkok-based software firm. He developed *Thai OCR*, and has managed many large-scale information retrieval and database projects, including the *Thai Democracy Project* (a record of all documents associated with the creation of the 1997 constitution), and the *Thai Legal Code Database*, as well as high-performance Web-based Java applications. Dr. Saipetch teaches mathematics and software engineering seminars at Mahidol University.

Paul Sidwell (Ph.D. U Melbourne) Dr. Sidwell is a Research Fellow in the *Department of Linguistics*, Australian National University. He has done wide-ranging fieldwork in Vietnam, and Laos, and has an extensive record of teaching and publication on Mon-Khmer languages, including five books, with a particular focus on the Bahnaric and Katuic language families. He serves as an editor of ANU's *Pacific Linguistics* publishing house, and established their XML metatagging standards for electronic dictionary submission and publication. Dr. Sidwell has founded or co-founded several cultural preservation projects in Laos, including the *Paksong Ethnic Literacy Promotion* and *Paksong Cultural Centre* projects. He is Director of Mon-Khmer projects for CRCL, and is currently leading the *Mon-Khmer Comparative Dictionary* project.

Justin Watkins (Ph.D. SOAS/U London) Dr. Watkins is a Lecturer in Burmese/Myanmar at the University of London *School of Oriental and African Studies*. His research interests include Burmese language and modern literature, descriptive and historical linguistics of Mon-Khmer and Tibeto-Burman languages, tone and register languages, experimental, acoustic and practical phonetics, computer lexicography, and literary translation. Dr. Watkins founded and directs the SOAS *Wa Dictionary & Internet Database for Minority Languages of Burma* project, which is compiling dictionaries, text corpora, and associated software.

APPENDIX 3. THE SEALANG LIBRARY

Applicant:

University of Wisconsin-Madison, Center for Southeast Asian Studies
 Address: 207 Ingraham Hall, 1155 Observatory Drive, Madison, WI 53706-1397
 Phone: 608 263-1755, Fax: 608 263-3735

Key Contacts:

Project Director: Prof. Robert J. Bickner, 1242 Van Hise Hall, University of Wisconsin-Madison, Madison, WI 53706. Phone: 608 262-3915, <rbickner@wisc.edu>

Project Technical Director: Doug Cooper, CRCL Inc., 820 Calle Pluma, San Clemente, CA 92673. <doug@th.net>

Project Dates: October 1, 2005 – September 30, 2009

Countries: Myanmar, Thailand, Lao PDR, Cambodia, Vietnam.

Languages: Burmese, Thai, Lao, Khmer, Vietnamese, Mon, Karen, Shan.

Cooperating Organizations:

Center for Southeast Asia Studies (University of Wisconsin-Madison, sponsor institution)

Center for Research in Computational Linguistics (CRCL Inc., co-sponsor)

Ecole française d'Extrême-Orient (EFEO, Burma, Thailand, Laos, Cambodia, Vietnam)

Center for Khmer Studies (CKS, Siem Reap)

Council of Teachers of Southeast Asian Languages (COTSEAL)

Committee on Resource Materials on Southeast Asia (CORMOSEA)

Southeast Asian NGO Projects: Thai Open Source (Thailand), **KhmerOS Project** (Cambodia), **Vietnamese Computational Linguistics Research Group** (Vietnam), **Lao Open Source** (Lao PDR), **Myanmar Unicode & NLP Research Center** (Myanmar)

Project Overview:

Modern Southeast Asia (SEA) includes eleven states, divided between the mainland countries Vietnam, Cambodia, Laos, Thailand, and Myanmar (population roughly 210 million), and the insular countries Malaysia, Indonesia, Brunei, Singapore, East Timor, and the Philippines. A global crossroads of singular geopolitical importance, Southeast Asia transits an astonishing 30% of the world's trade goods, including half its oil, through the central Malacca Straits.

Unfortunately, we have little capacity for information access or language reference in this important and often tumultuous region. The mainland countries are represented least: data and dictionaries that use the Indic-derived scripts of Burma, Laos, Thailand, Cambodia and the minority Mon, Karen, and Shan states, and Vietnam's Roman-derived, Chinese-influenced *Quốc Ngữ* script, are largely unavailable in the U.S., and there has been little software development beyond basic tools for text input and output.

The SEALang Library is a technically innovative plan to build core lexical and software resources for all Southeast Asian languages, starting with the difficult non-Roman scripts used in the five mainland countries. These will provide immediate access to a wide range of reference

and research functionality, and will help enable the development of more advanced lexical resources and information retrieval tools in the future.

The SEALang Library directly addresses TICFIA's goals:

- it produces the innovative technology required for information access and language instruction for writing systems that use non-Roman scripts;
- it has a regional focus, helping us understand the history of a strategically important area that suffers from deep political and cultural divides;
- it meets long-range teaching and scholarly needs, building software, reference, and research resources that will be of value for decades;
- it has broad value for the Southeast Asian Studies and heritage communities, and will enable research and development in a wide variety of specialties.

The SEALang Library will provide:

DICTIONARIES: we will prepare digital bilingual dictionaries with Web-based user interfaces, drawn from the best available scholarly print reference works – often difficult to obtain from U.S. libraries – for the national languages Burmese, Lao, Thai, Khmer, and Vietnamese, and for the major ethnic minority languages Mon, Karen, and Shan. Supplementary dictionaries will be added for languages (especially Lao and Khmer) that have undergone significant orthographic change in recent years. All dictionaries will be extended, where possible, by specialized lexicons of newly minted words, and all will incorporate the approximate search software tools described below.

The SEALang Library will add extensive XML metatagging to the source dictionaries to support data mining applications. All SEALang Library dictionaries will be accessible via approximate search software, and can be used both interactively, and as program-accessible Web resources. Queries may search for original orthography, transliteration, phonetic transcription, rhyme, part-of-speech, etymology, English content of definitions, and so on – if it's in the dictionary, it will have a specific search function. In the future, we anticipate that these bilingual dictionaries will be supplemented by monolingual references as well.

TEXT CORPORA: we will build monolingual and aligned bitext corpora. Used to study collocation and usage, and to support *data-driven* language learning, text corpora are necessary precursors to more advanced translation and monolingual and cross-language information retrieval tools. Teachers and students can query the corpora from the SEALang Library dictionaries (thus, a definition can be accompanied by common collocational contexts) or from other texts to supply authentic usage examples, lexicographers can find collocations or extract non-dictionary entries, and software developers and linguists alike can analyze word frequency or distributional data.

The SEALang Library will provide substantial (to tens of millions of words) monolingual corpora for the majority languages, along with the largest feasible (hundreds of thousands of words for Thai and Vietnamese, and less for others) aligned two-language corpora, drawn from both on-line resources and on-the-ground publishing contacts.

SOFTWARE: we will build information access tools for Southeast Asian scripts, including basic tools for segmentation and transliteration, conversion between font encodings, text harvesting and indexing, and statistical analysis. The user applications listed below will be linked to dictionaries, text and text corpora to help fulfill the promise of regional information access.

segmentation: the SEA-Seg utility will use dictionary data to segment vernacular texts into individual words.

cataloging: the SEA-Cat cataloging utility will help generate segmented, Library of Congress / ALA transliteration from vernacular orthographies.

query building: the SEA-Search searching utility will do the reverse, building local-orthography queries from transliteration or transcription.

text-as-image delivery: the SEA-See download utility will generate image data on the fly when Unicode and/or appropriate rendering software (a problem for Burmese and Khmer Unicode) are unavailable.

approximate dictionary search: utilities will provide intelligent heuristic-based approximate search capabilities for original orthography, transliteration, and/or phonetic transcription as available. These will dramatically improve the ease-of-use of the dictionaries.

corpus building and analysis: utilities will help collect, clean, and build the text corpora, align the bitext corpora, allow retrieval from very large unsegmented monolingual corpora, and analyze the corpora for data on word frequency and collocation.

transliteration, transcription, & conversion: utilities will orthographically transliterate or phonetically transcribe texts, and convert search items and older ‘legacy’ texts to and from Unicode, and the still widely used traditional font encodings, as possible.

The SEAlang Library is a long-awaited addition to the national digital infrastructure being built with the support of a variety of U.S. Department of Education Title VI programs. It will enable:

pedagogy and new teaching, learning, and translation tools for less-commonly taught languages,

scholarly inquiry in linguistics, history, lexicography/etymology, and Southeast Asia area studies,

scientific research in computational linguistics and cross-language information retrieval, and

language reference all but unavailable to 1.8 million Americans of mainland Southeast Asian heritage who can typically speak – but not read, or consult reference materials in – their heritage languages.

APPENDIX 4. EXAMPLES OF SEALANG LIBRARY PRELIMINARY DEVELOPMENT

The SEALang Library, funded by the Department of Education, and with cost-share funds provided by CRCL Inc., began its four-year development program on Oct. 1, 2005. Although this material is not scheduled to go on-line until early 2006, these are examples of preliminary development.

Phonetic (+ optional Thai) Haas Ratchabandit Both

A single period or whitespace **must** separate each syllable: **sa.baay.dii** or **sa baay dii**

Search options: ([details](#))

Complete head only (fewest hits).

Complete head, or syllable in a head.

Complete head, or head within a compound.

Syllable or longer (most hits).

Short rhymes (heads only) Long rhymes.

Doubles (see [details](#)).

Return search expansions only.

Approximate match options: ([details](#))

Vowel length and substitutions.

Leading consonants and clusters.

Common reading errors.

Regional accent variations.

Word-family match options: ([details](#))

/ / Variant initial / vowel / final consonant.

Any initial consonant(s) for diphthong.

Output options: ([details](#))

Order by corpus stats into synsets

Add Haas gloss to Ratch head words in defs

Allow Latin-name items from Ratch-only search

Allow odd syllable counts in doubles

Show Thai as Thai IPA Thai and IPA.

search			search		
IPA	Thai		IPA	Thai	
b	บ		C	ค ^h	คชณ
c/j	จ		K	ก ^h	กคณ
d	ด	ดฎ	N	ง	ง
f	ฟ	ฝ	P	พ ^h	ภพพ
h	ห	ฮ	T	ท ^h	ทธฎ
k	ก	ก			ทฬณ
l	ล	ลพ	a	า	-า
m	ม	ม	e	เ	-เ
n	น	ณ	i	ิ	-ิ
p	ป	ป	o	โ	-โ
r	ร	ร	u	ู	-ู
s	ส	ซสศษ	A	เอ	-เอ
t	ต	ตฎ	E	เ	-เ
w	ว	ว	O	อ	-อ
y	ย	ยญ	U	ุ	-ุ

Figure 1. Interface for Haas Thai-English / Royal Institute (Ratchabandit) Thai-Thai dictionary tool. We were primarily interested in exploring applications in linguistics and etymology in these tests, and put quite a bit of effort into exploring special-purpose queries. For example, the **Doubles** option finds constructions with internal rhyme or rhythm (like *helter-skelter* or *kith and kin*); these are common across Southeast Asia. In the *search/IPA/Thai* table on the right, note that in some cases, a single search letter (and sound) like **s** can have many Thai spellings (here, **s** maps to four Thai letters, while **T** – aspirated ‘t’ – has six). This can make using print dictionaries a tedious, discouraging experience for students.

Vowel+consonant near-misses (expanded to sa+[pbP]a+y): (3 more hits)

สบาย **sa·baay** 1 1 V to be well (in health), happy, comfortable. 2 (a place) to be comfortable.

สะพาย **sa·p^haay** 1 1 V to wear over the shoulder, carry on the shoulder.

สะใภ้ **sa·p^háy** 1 1 N [chiefly compounds] a female relative by marriage; designates a woman as an in-law of her husband's blood relative, and is typically combined with kinship terms. [opposite:0 เขย]

Figure 2. This shows the values returned for a Haas query. For English-speaking students, relatively subtle variations in consonant sound (especially between Thai **b-p-ph**) and vowel length (esp. **a-aa**) can make it very difficult to guess at proper spellings. This kind of ‘intelligent’ wild-card – limited to the plausible – maximizes the student’s chance of finding the right word on the first try.

Expansion of: sa.bay

Vowel near-misses (expanded to sa+.ba+y).

Consonant near-misses (expanded to sa.[pbP]ay).

Vowel+consonant near-misses (expanded to sa+[pbP]a+y).

Reading errors involving double-functioning consonants and consonant clusters (expanded to sa.bay).

Reading errors + vowel near-misses (expanded to sa+.ba+y).

Common regional accent substitutions (expanded to Ca.[pbP][lr]?ay).

Accent substitutions + vowel near-misses (expanded to Ca+[pbP][lr]?a+y).

Historical initial consonant alterations (expanded to [sjC]a+[mbpP]a+y).

Historical final consonant alterations (expanded to sa.ba[nty]).

Historical vowel alterations (expanded to s[UAa]+a?.b[UAa]+a?y).

Historical vowel+consonant alterations (expanded to s[UAa]+a?.b[UAa]+a?[nty]).

Historical initial consonant + final consonant alterations (expanded to [sjC]a+[mbpP]a+[nty]).

Historical consonant+vowel+consonant alterations (expanded to [sjC][UAa]+a?.[mbpP][UAa]+a?[nty]).

Rhymes (expanded to [bcdfhijklmnrstwyPTKCN]*a.bay).

Key:

[abc] Any one of a, b, or c.

x? Zero or one of x (thus [abc]? is zero or one of a, b, or c.)

x+ One or more of x (thus [abc]+ is one or more of a, b, or c.)

x* Zero or more of x (thus [abc]* is zero or more of a, b, or c.)

Figure 3. This capture of the *return search expansions only* option provide a peek at how we develop ‘intelligent’ search heuristics. All of these search expansions are limited to plausible alternatives, based either on perceptual errors we can anticipate a student making, or on regular historical or regional sound change. Tools like this have a wide range of applications, helping to solve problems from etymology to spell checking.

LC/IALA á à â ã ä å ē ī ū ý ŷ ʀ ʁ ʌ ɪ

IPA ɔ ɛ ə ɯ ʊ ɔ ɑ ɲ ɳ ʒ

TEXT POS Usage Subject Etymology Dialect

heads only ignore LC diacritics

LC / IPA: V = any vowel / D = any diphthong / C = any consonant / double = match all doubles
 TEXT: 123 = head 123 / p123 = page 123 / 123.4 = page 123, item 4

Figure 4. A different approach to user interface design; this targets Headley’s Cambodian-English Dictionary. It allows search by Library of Congress transliteration (top two rows of buttons) or International Phonetic Alphabet transcription (third row). The **V***, **V+**, **C+**, and **CV+** buttons show a simplified approach (in comparison to our Thai tool in Figure 1) for managing vowel or consonant approximation. The **POS**, **Usage**, and other drop-downs both refine searches, and allow data mining operations. They rely on extensive XML metatagging in the underlying text.

Return contexts collocates
Restrict colls to nouns verbs numbers

A = B Compare usage of words **A** and **B**.
A ~ B Find **A** if it is near **B** (use **A ~~ B** to allow **A ~ B** or **B ~ A**).

Search returns contexts or collocates from the selected corpus.
I feel lazy selects largest corpus <= target size if possible (indexed words only).

-hit target for 'feeling lazy.'

top items with stats.

-character context before and after.

-char max gap between *near* terms.

-hit minimum before printing contexts.

first instances of any context.

More about [texts](#) [search types](#)
[statistical ranking](#) [match restrictions](#)
[output restrictions](#) [segmenting strategy](#)

- Literary text (.21M)
- Law codes (20M)
- Tax code (1.3M)
- ORCHID corpus (3.9M)
- Thai News Agency (1994-96) (49M)
- Royal-related news (94-96) (1M)
- Kanchanpisek set (8.6M)
- Newspaper set (47.5M)
- Small Web set (18.6M)
- Big Web set (140M)
-

- Sort (and stats) by leading collocates.
- Sort by center collocates (stats tie to first string second string).
- Sort (and stats) by trailing collocates.

- Return co-occurrence statistics (rank by: log-likelihood mutual information t-score frequency Dice co-ef)
- No stats (sort by ASCII order of collocates).

- Ignore search term when found within a: headword compound word.
- Allow even obvious partial-word matches.
- Allow even obvious mid-syllable matches.

- Return statistics only.
- Don't format text, do bracket {targets}.

Figure 5. A preliminary Thai text corpus tool. Such tools have never been built for working with nonsegmented alphabetical texts, which pose interesting analytical problems because all statistical methods require accurate word counts. This tool allows localized segmentation of the target's immediate environment, and lets us test various statistical approaches to analyzing collocation when exact word counts are uncertain. It also begins to explore characteristic formations, such as 'reversible' doubles; these are unusual in English – e.g. **skelter-helter* – but are common in Southeast Asia.

บ้าน: 5332 matches (from smallweb); Highlighting non-Haas items that appear at least 5 times. Underlining collocates that appear left and right.

[Study contexts](#) Highlight / doubleclick a phrase; click to show contexts (right for leading, left for trailing).

552 distinct leading collocates:	920 distinct trailing collocates:
() บ้าน (403) ในบ้าน sV (391) ที่บ้าน	(306) <u>บ้านเรา</u> P (123) บ้านเมือง N
N, C, Pp, rP (293) <u>เรื่องบ้าน</u> N, C, Pp	(108) <u>บ้านที่</u> N, C, Pp, rP (106) <u>บ้านเลข</u> Cj
(230) <u>กลับบ้าน</u> V, sV, AA	(89) <u>บ้านหรือ</u> Cj, Pt (83) <u>บ้านก็</u> AA, Pt
(139) <u>นอกบ้าน</u> sV	(80) <u>บ้านของ</u> N, Pp (59) <u>บ้านได้</u> V, AA, sV
===== 25% (top 6 items = 1456	(59) <u>บ้านคุณ</u> N, P, T (55) <u>บ้านผม</u> N, P
of 5332 hits)	(50) <u>บ้านให้</u> V, sV (49) <u>บ้านอยู่</u> V, sV
(130) หมู่บ้าน C (119) เพื่อนบ้าน N	(48) <u>บ้านหลัง</u> N, C, Pp (48) <u>บ้านไป</u> V, sV
(102) อยู่บ้าน V, sV (101) <u>จากบ้าน</u> N,	(47) บ้านเรือน N, C (46) บ้านพัก V, C
V, Pp (67) แม่บ้าน N, P, T	===== 25% (top 16 items = 1356 of
(65) <u>ขายบ้าน</u> V (60) <u>ของบ้าน</u> N, Pp	5332 hits)
(57) <u>พื้นบ้าน</u> N (56) <u>หน้าบ้าน</u> N, Pp, C	(41) <u>บ้านมี</u> จุดด้วยกัน, V (40) <u>บ้านเป็น</u> V, sV

Figure 6. Leading and trailing collocates of the Thai word for ‘house.’ Clicking any item shows all the contexts it occurs in. The underlined items appear as both **target + X** and **X + target**. This particular application helps reveal important prepositional constructs (that are not typically noted in dictionaries), as well as important secondary meanings found only in compound constructions (e.g. ‘domestic’ and ‘community’).

Thai On-Line Library - Bitext Corpus
 Maintained by Doug Cooper (bugs to doug@th.net)
 Center for Research in Computational Linguistics, Bangkok <http://crcl.th.net>

[ABOUT THE BITEXT CORPUS](#) [WHY A BITEXT CORPUS?](#) [IMPLEMENTATION DETAILS](#) [RESEARCH PROJECTS](#)

About the Thai Bitext Corpus
 The Thai Bitext Corpus is a collection of Thai and (mostly) English parallel translations or *bitexts*. The complete library can be searched for usage examples, or individual texts can be read in a variety of layouts.

Bitext searches allow either Thai or any available second language (L2), and use an extended AltaVista 'advanced match' syntax: **not near and or** (! ~ & |), and parentheses: A **and** (B **or** C).

Thai: near = 20 chars

Search Reset L2: English / match case / full word / near = 10 words

Title	About the author and translator	Side-by-side lines		Over-and-under lines	
		paragraphs	paragraphs	paragraphs	paragraphs
The Muddy Path	Paitoon Thanya (tr. Tom Glass)	Go	Go	Go	Go
Building Sandpiles	Paitoon Thanya (tr. Tom Glass)	Go	Go	Go	Go
The Man and the Tiger	Seksan Presertkul (tr. Marcel Barang)	Eng. Fr.	Eng. Fr.	Eng. Fr.	Eng. Fr.
the story of jan darra (pp 77-87)	Utsana Phleungtham (tr. Pongdej Jiengpattanakij)	Go	Go	Go	Go
mad dogs & co. (pp 13-29)	Chart Korbijitti (tr. Marcel Barang)	Go	Go	Go	Go
The Wizard of Oz	L. Frank Baum (Thai Charnvit Kasetsiri, Lat. Hinke / Van Buren)	Eng. Lat.	Eng. Lat.	Eng. Lat.	Eng. Lat.

All texts are copyrighted by author, translator, and/or publisher, and are reprinted here by permission.

Figure 7. This bitext corpus is a limited on-line resource, designed to explore alternative approaches to showing aligned texts, and to explore problems of actually doing the alignments. It includes aligned texts for Thai-English, Thai-French, and Thai-Latin. It is the only such on-line resource for any mainland SEA language.

Search Results (Thai-English bitexts) for บ้าน: 133 lines, returning shortest items first (duplicate lines suppressed). Choosing 50 items at random.

This is my house!" [Building]	นี่บ้านฉัน"
What house? [Building]	บ้านอะไร
"The house must have fallen on her. [Wonderful]	บ้านคงตกลงมาทับเธอแน่ๆ
And you, spirits of all ilk, help me! [story]	ที่บ้านผีเรือน !...
There is no place like home." [Wonderful]	ไม่มีที่ไหนเหมือนบ้านหรอก
I'm so glad to be at home again!" [Wonderful]	หนูดีใจจังที่ได้กลับบ้านอีก!"
See!" she continued, pointing to the corner of the house. [Wonderful]	ดูซิ"เธอล่าวต่อแล้วชี้ไปที่มุมบ้าน

Figure 8. Bitext hits for the Thai word for ‘house.’ This sort of reference, using both English and Thai-original texts, is enormously helpful for receptive and productive translation. And like the single-language corpus tool, it can provide an important reference resource for correctly formulating search queries.

Patruus Henericus et Amita Em magnum lectum in alio angulo, et Dorothea lectulum in alio habebat.	ลุงเฮนริกกับป้าเอมีมีเตียงนอนใหญ่อยู่ที่มุมหนึ่ง ส่วนโด โรธีมีเตียงเล็กอีกที่มุมหนึ่ง
Nullum omnino cenaculum erat, neque cellanisi foramen parvum, in terra fossum, quod vocatur cella turbida (id est refugium a tempestatibus), ubi familia descendere poterat si casu unus ex illis magnis turbinibus, tam validus ut omne aedificium quod obstabat opprimere posset, oreretur.	ไม่มีห้องใต้ดินเลย ห้องใต้ดินก็ไม่มีเว้นแต่มีโพรงเล็กๆที่ขุดไปได้พื้น เรียกว่า โพรงไซโคลนเป็นที่ครอบครวันี้จะมีมุดเข้าไปเมื่อเกิดลมพายุซึ่งกระโชกแรงจนบดขยี้สิ่งก่อสร้างใดๆที่ขวางทางมันได้

Figure 9. A sample of the Latin-Thai *Wizard of Oz*. Because Thai includes a considerable number of Indo-European loans (via Indic), there are a surprising number of cognates in the text. This is common across Southeast Asia, yet there are no references available to help language learners.

APPENDIX 5. JAWI SCRIPT

Below, selections from two articles (**5a**) *Script and Identity in Southeast Asia*, van Dijk, and **5b**) *Digitisation of an Endangered Written Language: the Case of the Jawi Script*, Yaacob et al.) Both refer to the use of Jawi in Southern Thailand. Note van Dijk's comments in particular:

Data from Singapore and South Thailand indicate that the desire to study Arabic was further stimulated in the aftermath of September 11 ... In South Thailand and the southern Philippines, Jawi has assumed extra significance as the script of a minority Muslim population. In both regions the Muslim inhabitants feel discriminated against and perceive their own culture to be under threat. In both regions past instances of suppression and attempts to have them assimilate into the dominant culture have triggered separatist movements.

5a) From *Script and identity in Southeast Asia*, Kees van Dijk, International Institute for Asian Studies newsletter, 37, July 2005 <http://www.ias.nl/ias/show/id=49979/framenoid=42832>

Script Script also enters the picture. In Southeast Asia one of the choices in colonial times, which still persists today, is that between Roman and Arabic script. If choice of language evokes strong emotions, the debate about script introduces a new, sensitive dimension: religion. In the Netherlands Indies there was a tendency among Indonesians and Europeans alike to associate Malay in Arabic characters with Islam and its expansion, while Muslims rejected Roman script because of its association with Christianity. From their side, some Dutchmen presented the promotion of Roman script as the most salutary remedy to halt the advance of Islam.



In the end Roman script became the accepted script for the public sphere. Jawi, as it were, became the script of religious experts. In Malaysia this happened only in the 1950s-60s, much later than in Indonesia. Nevertheless, knowledge and use of Jawi script may well be on the rise, a trend partly brought about by a factor at work all over Southeast Asia. The revival of Islam has resulted in more people wanting to be able to read and write in Arabic, whether 'pure' Arabic or Jawi and Pegon (Arabic script adjusted to Malay, Sundanese, and Javanese).

Data from Singapore and South Thailand indicate that the desire to study Arabic was further stimulated in the aftermath of September 11. Circumstances in individual countries also contribute to the wider trend: the Malaysian government, for example, now pays greater attention to religious education - including the teaching of Jawi at public primary schools - to create an alternative to a specific type of Islamic school, the Sekolah Agama Rakyat (SAR) or People's Religious School, seen as the bulwark of a radical opposition.

In South Thailand and the southern Philippines, Jawi has assumed extra significance as the script of a minority Muslim population. In both regions the Muslim inhabitants feel discriminated against and perceive their own culture to be under threat. In both regions past instances of suppression and attempts to have them assimilate into the dominant culture have triggered separatist movements.

The Indonesian government during the New Order embarked on a similar policy that aimed at a complete ban on the public manifestation of Chinese culture, which included prohibition on the publication of Chinese books, newspapers (except for one which was controlled by the intelligence agency BAKIN) and magazines. The ban was part of a wider campaign spearheaded by the military to put an end to public display of Chinese characters, a policy which was discontinued only after the fall of Soeharto.

If the use of Arabic and Jawi were fueled by religious developments, economic factors are at work in the contemporary promotion of Chinese language and script. Proficiency in the latter is presented not only as markers of identity - crucial aspects of one's civilization and indispensable links to one's cultural heritage. The growing economic might of the People's Republic of China and expectations of the Chinese economy's importance in the world also play a role.

Though non-Christians in South and Southeast Asia in the colonial era may have loathed the West for its perceived immorality and permissiveness, European culture's associations with modernity, emancipation and power held great appeal for many. Today global countervailing forces are stronger, making it easier for cultural traits associated with non-Western cultures and non-Christian religions not only to maintain their position, but to regain some ground.

Kees van Dijk is affiliated with the KITLV/Royal Netherlands Institute of Southeast Asian and Caribbean Studies and is professor of history of Islam in Indonesia at Leiden University. He is interested in contemporary political developments in Southeast Asia and in the cultural manifestations of nation-building and community identity.

5b) From *DIGITISATION OF AN ENDANGERED WRITTEN LANGUAGE: THE CASE OF THE JAWI SCRIPT*** Mashkuri Yaacob; Zainab A.N.; Rohana Mahmud & Nor Edzan Che Nasir, University of Malaya, Kuala Lumpur.**
http://www.unesco.or.kr/cyberlang/data/Mashkuri_Yaacob.doc

The Malay language is Malaysia's national language and is the language of the Malays of the Malay Archipelago (Malaysia, Singapore, Brunei, Southern Thailand and Indonesia). The other spoken languages, however has their roots from the languages in India and China.

The Malay language was the spoken word before the written script evolved. Jawi originates from the Arabic scripts and evolved in the Malay world as early as 440H (1104A.D.) (Nafisah, 1999). An evidence of this is the inscription found on a tombstone dated 1303 AD in Terengganu. As such, the early written script for the Malay language naturally evolved from the Arabic alphabets. The Jawi alphabets comprise 29 Arabic characters and 6 additional letters devised by the Malays to accommodate local vocal sounds. The British invented the romanised Malay script when they colonised the Malay Peninsular in the 18th century and the English language had subsequently greatly influenced the spelling structure of the Malay language until it was standardised in the post 1973 years.

The romanised Malay or Rumi survived and thrived. The use of Jawi Malay scripts however, dwindled in an alarming rate and is considered to be an endangered script. It was once widely used in the Malay courts and was the dominant writing in the Malay world. Today the Jawi script is mainly used for Islamic religious documents and texts. There are factors that threatened even these usages of Jawi. One dominant factor is the growing reluctance among local publishers to publish religious books for the public mainly in Jawi. Economic pressures favour publications in Rumi Malay, that would include quotations from the Quran printed in Arabic script. This is because fewer Malays are Jawi literate and the situation is exacerbated by the wide availability of romanised word processing software that can easily accommodate Rumi Malay (Muhammad Mun'im and Haliza, 1994). The consequences was felt by the national Jawi daily newspaper *Utusan Melayu*, which almost stopped its print run due to lack of sales (Ahmad Zaki, 1998).

REVIVALISM THROUGH DIGITISATION

An attempt to revive the use of Jawi was initialised when the Ministry of Education in Malaysia introduced the teaching of Jawi script in public primary schools. The essence of this is reflected in one of the mission statement of the Department of Islamic and Moral Education Malaysia, that is, to ensure that every Muslim child that completes year six at the primary level can read the Quran as well as read and write Jawi. This skill is taught within the Islamic education curriculum. Even so, the rate of Jawi literate Malay children remained low today and there is growing concern that the future generation would not be able to read the literary text of their national heritage (*Study on learning...*, 1989). The study found that among the sample of 853 standard six students from 24 schools, throughout Malaysia, only 68.23% could read Jawi and 58.34% of the students could write it.

Another approach is to popularize the use of Jawi through information technology. Research and development in this area was dominated in the early years by academics from Universiti Teknologi Malaysia (UTM). UTM produced the first prototype computer, which can handle the Jawi script, in 1983 (Ahmad Zaki, (1986, 1987, 1998). UTM also collaborated with the Standards and Industrial Research Institute of Malaysia (SIRIM) to devise a standard for Jawi national character set for data interchange purposes. This led to the development of a character set compatible with the ISO8859 Arabic character set. The Jawi software developed could handle left-to-right and right-to-left script writings. UTM also designed a new keyboard layout that supported Jawi character input. The software and keyboard is used by UTM to enter data into their *Quran Information System* (QIS) project (Mohd. Shazali, 1990). Today, those working with Jawi text use custom fonts either in the Windows or Macintosh environment. The Jawi provided by most software is a modification of the Arabic font. Examples of such software are *WinText* and *AlKaatib*. Also, the Jawi characters are based on the ISO 9036 code set (ISO9036, 1987) which defines the stand-alone version of Arabic characters in a form that can be used for interchange between computer systems using a 7-bit code set. This is supplemented by the ISO 1182 (1996) that defines the use of Arabic alphabet character set for bibliographic information interchange. UTM and SIRIM were also active in the area of automatic conversion of Rumi text into Jawi. The method used was to break up the Rumi phonemes and map it to the corresponding phonemes in Jawi. There are numerous problems that need to be surmounted in this context, because the multiplicity of mapping may result in errors, especially when handling words which have multiple meanings or variant pronunciations (Muhammad Mun'im, 1994).

The Malaysian government's involvement in popularizing the use of Jawi was more active through the Department of Islamic Development Malaysia (JAKIM), which was established by

the Malaysian Council of Rulers (Majlis Raja-Raja Malaysia) in January 1997. One of JAKIM's mission is to increase the reverence and acceptance of the Jawi script by using information technology as an enabler (*JAKIM*, 2001). A memorandum of understanding was signed between JAKIM, the Association of Jawi Writing Enthusiasts (PENJAWIM) and a software company (Allis Tech) in 1997 and as a result, the JAWINET homepage was launched and was maintained by PENJAWIM. The project enables selected schools with computers that can support and have Internet connections send and receive e-mails, as well as post web pages in Jawi. The project uses a multilingual browser called *Tango* distributed by Allis Technologies Inc

To activate more participation from the educational institutions, another MOU was signed between JAKIM, PENJAWIM and the University of Malaya in March 1998. In the same year, a workshop on JAWINET was held at the Faculty of Computer Science and Information Technology (FCSIT), University of Malaya, which brought together parties interested in popularizing the Jawi script. FCSIT subsequently continued the JAWINET project and this extended into the *Digital Jawi Project (DJP)*.

The focus activity of the collaborative parties is on research and development and dissemination of Jawi applications especially through the Internet. The JAKIM, PENJAWIM and UM collaboration have five main issues to tackle:

- (a) To produce the architecture, standard, technology, product and services related to digital Jawi and for the use by JAWINET users;
- (b) To produce an Internet software for the JAWINET system to enable Internet users to surf, read, author and communicate in Rumi or Jawi Malay;
- (c) To organise educational and training programmes to encourage the use of digital Jawi technology;
- (d) To form an association that gather members from various fields who could contribute to the continuous development of the digital Jawi; and
- (e) To promote awareness about digital Jawi and JAWINET (Ahmad Zaki, 1998).

The University of Malaya currently provides and maintain a Digital Jawi Laboratory, where most of the Jawi script projects were developed. Collaborative efforts were also geared towards improving the JAWINET homepage that disseminates information about research and development as well as activities of the Digital Jawi Project.

The research activities on Jawi have spread to other local universities. Khairuddin and Ramlan (1996) first mentioned the application of neural network (NN) techniques at a conference in Serdang, Malaysia. The NN technique was used to classify Jawi characters. Following this, researchers at the Mara University of Technology (UiTM) and National University of Malaysia (UKM) collaborated on the application of recurrent neural network techniques in recognizing handwritten Jawi words. Work on this project was first reported in 1998 (Mazani, et al.) and after four years, the researchers revealed that recurrent neural networks can be applied to solve handwritten recognition problems (Mazani, et al, 2001). This result means that it would be possible for the Jawi illiterate reader to understand old literary handwritten texts. Although this goes against the grain of encouraging researchers to read Jawi, it helps promote research on Malay manuscripts. At the Multimedia University three researchers are developing a teaching software that aids users in the learning of Arabic calligraphy. The software can be used to test handwritten Arabic characters for correctness specific to the Thuluth calligraphy method (Nor

Rafeah, Seyed Mohamed and Akbar, 2001). The Universiti Sans Malaysia (USM) and Winsoft of France have produced a Jawi word processing programme called *Winsoft Jawi* in 1994 (UTM, 2001).

REFERENCES

- Ahmad Zaki Abu Bakar. 1998. Preserving a national heritage through multilingual information technology: country report. *MLIT Symposium: GII/GIS for Equal Language Opportunity, 6-7 October 1998, Ha Noi, Vietnam*: 4p. Also available at: <http://www.cicc.or.jp/english/hyoujyunka/mlit3/7-4.html>.
- Ahmad Zaki Abu Bakar. 1986. Natural language processing of Jawi script. Paper presented at the first *JSPS-VCC Seminar on Integrated Engineering* held on 11-12 December 1986 in Kuala Lumpur. Kyoto: Faculty of Engineering, Kyoto University: 343-346.
- Ahmad Zaki Abu Bakar. Natural language processing and understanding of bahasa Malaysia. Paper presented at the *third Malaysian National Computer Conference*, held at 19-20 August 1987 in Kuala Lumpur.
- Gordon, Peter. 1990. Development of a Rumi to Jawi translator. Paper presented at the *International Conference on Information Technology*, 17-20 September 1990, Kuala Lumpur: 11p.
- Mohd. Shazali. 1990. Databases in Malaysia: country report. Paper presented at the 4th *AFSIT, October 24, 1990, Tokyo*. Available at: <http://www.cicc.or.jp/english/hyoujyunka/af04/4-06.html>
- ISO 9036, Available at: http://www.dkuug.dk/il8n/charmmaps/ASMO_449.
- JAKIM*. Available at: <http://www.islam.gov.my>
- Khairuddin Omar and Ramlan Mahmud. 1996. Genetik-rangkaian neural untuk pengkelasan aksara Jawi. Paper presented at the *National Conference on Research and Development in Computer Science and its Applications*, 26-27 June 1996, Serdang:p.5-9.
- Malaysia's data on population. 2000. Available at: <http://www.ids.org.my/stats/KeyData/population.htm>
- Mazani Manaf, Mohd Jamil Abu Sari; Abdul Razak Hamdan and Muhammad Mohd Yusof. 1998. Recognition of handwritten Jawi script and transliteration of Jawi scripts to Rumi scripts. Paper presented at the *International Conference on Information Technology and Multimedia*, 28-30 September 1998, Kajang: 7p.
- Mazani Manaf, Mohd Jamil Abu Sari; Abdul Razak Hamdan & Muhammad Mohd Yusof. 2001. Application of Bama optimised recurrent neural network in handwritten recognition of Jawi works. Paper resented at the *International Conference of Information Technology and Multimedia*, 13-15 August, Universiti Tenaga Nasional, Bangi.
- Muhammad Munim Ahmad Zabidi and Haliza Ibrahim. 1994. Text processing in Malay, *AFSIT*, Tokyo, 27th September: 9p. Also available at: <http://www.cicc.or.jp/english/hyoujyunka/af08/8-07.html>.
- Nafisah Ahmad. 1999. Romanization of multiscript/multilingual materials: experiences of Malaysia. 65th IFLA Annual Conference, 20-28 August 1999, Bangkok.
- Nor Rofeah A. Sani and Seyed Mohamed Buhari. 2001. A Method of recognition of Arabic calligraphy characters. Paper presented at the *International Conference of Information Technology and Multimedia*, 13-15 August, Universiti Tenaga Nasional, Bangi.
- A study on learning Islamic education in the primary schools in Peninsular Malaysia*. 1989. Kuala Lumpur: Educational Planning and Research Division (EPRD). Available at: <http://www2.moe.gov.my/~bppdp/abs/1989.htm>.

APPENDIX 6. WEB ACCESSIBILITY INITIATIVE

*The Digital Dictionaries of South Asia project and Digital South Asia Library have prepared the following statement describing their intent to strive for Level II compliance with the guidelines of the **Web Accessibility Initiative**. The Digital Southeast Asia Library also intend to follow these principles (statement taken from dsal.uchicago.edu/techinfo.html)*

Web Accessibility Initiative

Introduction

In 1998, Congress amended the Rehabilitation Act to require Federal agencies to make their electronic and information technology accessible to the disabled. Recent legal interpretations of the act have extended the scope of the law to all state-controlled colleges and universities. The criteria for compliance with regard to Web-based technology and information are based on access guidelines developed by the Web Accessibility Initiative of the World Wide Web Consortium (W3C). These standards consider the needs of users with a variety of disabilities. People with visual disabilities will find it easier to use reading browsers because of the labeled graphics, described video, marked up tables, and guidelines for the use of color and movement. People with hearing disabilities will benefit from the captions included with audio files. Those with physical disabilities that limit their capacity to use a mouse can more easily navigate Web sites by means of keyboard and/or single-switch support for menu commands. Because users with cognitive or neurological disabilities often need a more consistent structure of information, WAI includes recommendations about data structure and display such as consistent navigation, concise language, and the elimination of flickering tags.

The World Wide Web Consortium has identified three levels of compliance with WAI. Level One compliance will remove the major barriers for users with specific disabilities. Levels Two and Three will incrementally improve the transfer of information. DSAL and DDSA Web pages will strive for Level Two compliance

There are coincidental benefits of the WAI for Web users without disabilities. For example, the various modalities of the WAI offer text only options for the users of Web phones or palm pilots with either small or even text only display screens. Clear and concise Web pages are beneficial to all users.

Specifics

After reading and reviewing a number of articles on the WAI, DSAL and DDSA have put together a brief list of the main points of WAI that affect our site.

HTML should help to convey meaning through its layout and presentation.

- The layout should be consistent throughout an entire site.
- Documents should be organized so that they can be read without style sheets. For example, when an HTML document is rendered without associated style sheets, it must still be possible to read the document.

Color.

- All information conveyed with color should also be possible to interpret without color, either from context or markup.

- Color blind individuals might not distinguish red from green from yellow. If you use color in Web design, use clearly contrasting colors such as blue and black and white and yellow.
- Eschew instructions based on color, such as 'push the red button'.

Fonts.

- Choose tags that emphasize translatable structure. Speech synthesizers are unable to interpret some forms of visual representation (such as 'italic'). Instead, use tags such as STRONG, EM, and H1 to add structure to documents.
- Avoid deprecated, that is to say outdated, font tags, for example (font color="blue")
- Separate navigation items with a space, a "|", and a space in order to avoid overtaxing screen readers.

Hypertext.

- Use phrases that are comprehensible out of context. For example, avoid "click here."
- Link text should be concise.
- Added access keys may help users without mice and work well with the audio browsers needed by some disabled users. For example, if a link entitled "To the Objectives Page" were included, users might type or speak "O" rather than clicking upon the link to type or speak "O" instead of clicking the "To the Objectives Page" link.

Tables.

- Explain table contents by use of the summary attribute in the table tag.
- Make line by line reading comprehensible.
- Use the summary tag for graphs and charts as well.

Declare the language of the document, for example (HTML lang="eng").

- This helps the computer or assistive device present information in a manner appropriate to that language and also helps software that translates text from one language into another.
- Use three letter ISO 639-2 language codes rather than two letter codes. The ISO list is available at <http://lweb.loc.gov/standards/iso639-2/englangn.html>

Avoid using frames as these cause problems for indexing and searching via search engines because of the way in which some spiders work.

Every image, sound, applet, and script should be accompanied by alternative textual descriptions.

- For example, for an image used to describe membership in a document, the gif file referring to membership information should be accompanied with a textual description in following manner, `img src="littlemembership.gif" alt="Membership Information" border="0"`
- Other options for text accompaniments, such as D-links or longdesc. are not currently recommended.

Provide alternative content for scripts, applets, and plug-ins in case active features are inaccessible or unsupported.

Avoid tags that cause screen flicker. Flickering can initiate seizures or other unintended consequences for some users.

Works Referenced

- Web Accessibility Initiative (WAI) from W3C <http://www.w3.org/WAI/>.
- Checklist for complete instructions <http://www.w3.org/TR/WCAG10/full-checklist.html>.
- Federal Information Technology Accessibility Initiative <http://www.section508.gov/> .
- Designing for the Color-Challenged: A Challenge http://www.InternetTG.org/newsletter/mar99/accessibility_color_challenged.html.
- Web Monkey <http://hotwired.lycos.com/webmonkey/97/11/index4a.html>, and http://hotwired.lycos.com/webmonkey/design/site_building/tutorials/tutorial5.html.

Tools

- <http://www.delorie.com/web/lynxview.html> Shows you what a given page would look like in Lynx, a text only browser. This can be a good place to start checking for compliance with WAI.
- <http://www.delorie.com/web/purify.html> Check the level of HTML compliance.
- <http://www.cast.org/bobby/> this tool will check one page at a time for the various levels of compliance (1 - 3).
- <http://watson.addy.com/> Dr. Watson is a free service that estimates download times and checks link validity, site popularity, and spelling.

APPENDIX 7. COOPERATING INSTITUTIONS

All information here is adapted from on-line sources

Foreign Service Institute (FSI)

The Foreign Service Institute is the Federal Government's primary training institution for officers and support personnel of the U.S. foreign affairs community, preparing American diplomats and other professionals to advance U.S. foreign affairs interests overseas and in Washington. At the George P. Shultz National Foreign Affairs Training Center, the FSI provides more than 450 courses – including some 70 foreign languages – to more than 50,000 enrollees a year from the State Department and more than 40 other government agencies and the military service branches.

The Institute's programs include training for the professional development of Foreign Service administrative, consular, economic/commercial, political, and public diplomacy officers; for specialists in the fields of information management, office management, security, and medical practitioners and nurses; for Foreign Service Nationals who work at U.S. posts around the world; and for Civil Service employees of the State Department and other agencies.

Ranging in length from a half-day to 2 years, courses are designed to promote successful performance in each professional assignment, to ease the adjustment to other countries and cultures, and to enhance the leadership and management capabilities of the U.S. foreign affairs community. Other courses and services help family members prepare for the demands of a mobile lifestyle and living abroad. FSI has one program available to the general public. Recently FSI has worked with the Overseas Security Advisory Council (OSAC) to make security awareness training available to U.S. private business members that operate overseas. This training is provided on a reimbursable basis.

The Foreign Service Institute has developed a series of language-learning courses for use by the state department and other government agencies. These courses have been developed, tested by use, and improved over a period of many years. Working with the Departments of Defense, Education, other government agencies, and various universities, the FSI has created the largest body of complete courses for languages of the world available anywhere.

The courses consist of a large book prepared by experts in the language, and a number of tapes or CDs recorded by native speakers. Many languages are presented in levels (up to four), so students can go on to achieve greater mastery of the languages they choose. The first level of these courses is sufficient for a working knowledge of the grammar and pronunciation, and will serve the needs of travelers and business people. All of these courses present the subject language in a way that allows students to master the material quickly and thoroughly, while moving at their own pace through lessons and reviews.

Defense Language Institute Foreign Language Center (DLI/FLC)

The Defense Language Institute Foreign Language Center (DLIFLC) is the world's largest foreign language institute. The mission of the DLIFLC is to educate, sustain, evaluate, and support foreign language specialists under the guidelines of the Defense Foreign Language Program, which provides the Department of Defense and other Federal agencies with linguists able to support United States national interests worldwide. To accomplish its four primary mission area and ensure that our military forces are prepared to meet global foreign language requirements, the DLIFLC remains flexible and responsive to the changing world order and on

the cutting edge of technological innovations in foreign language education for the twenty-first century.

The DLIFLC is regarded as one of the finest schools for foreign language instruction in the world. Resident instruction is provided at the Presidio of Monterey in some 23 languages plus several dialects and is administered through 31 language departments and the Operation Enduring Freedom Task Force, which expands or contracts in response to the needs of the sponsoring agencies. The present facilities at the Presidio of Monterey can accommodate approximately 3,500 students. Instruction is also routinely provided under DLIFLC-supervised contractual arrangements in Washington, D.C., in over 65 languages and dialects. The DLIFLC also provides extensive nonresident instructional support in a variety of languages and dialects.

All basic foreign language teaching takes place in one of eight Schools, plus the Emerging Languages Task Force (ELTF): Asian I (Chinese Mandarin, Japanese, Tagalog, and Thai); Asian II (Korean); the Multilanguage School (Serbian/Croatian, Persian-Farsi, Dari, Hebrew, Turkish, Italian, French, German and Portuguese), the European and Latin American School (Russian and Spanish) Middle East I (Arabic); Middle East II (Arabic); Middle East III (Arabic), the Continuing Education Directorate and ELTF (Pashto, Kurdish, Hindi, Uzbek and Urdu).

Each School is headed by a civilian Dean responsible for planning and implementing assigned programs in foreign language training and curriculum development, implementing academic and administrative policy, and managing the School's annual manpower and budget allocations. An Associate Dean, who is a senior military officer, provides counsel and assistance to the Dean, monitors student progress, and directs the School's Military Language Instructor Program.

Each school is composed of departments, in which instruction of individual foreign languages takes place. Each department is headed by a civilian chairperson, who is responsible for the instructional program, manages the assigned instructors and staff, and oversees foreign language education and the faculty development process. Instructors, organized into teams, are responsible for teaching classes, evaluating student performance, and developing and maintaining course materials.

The Defense Language Institute Foreign Language Center is accredited by the Accrediting Commission for Community and Junior Colleges of the Western Association of Schools and Colleges, an institutional accrediting body recognized by the Council on Higher Education Accreditation (CHEA) and the U.S. Department of Education.

The Defense Language Institute Foreign Language Center's efforts in achieving academic excellence were recognized in 1979, when the Accrediting Commission for Community and Junior Colleges (ACCJC) of the Western Association of Schools and Colleges granted it accreditation. This status was last reaffirmed in 2000. The DLIFLC falls under the jurisdiction of the ACCJC as a specialized post secondary institution. In October 2001, the United States House of Representatives approved DLIFLC's request for Federal degree-granting authority. Students who meet the graduation requirements are awarded an Associate of Arts degree from DLIFLC. In addition, DLIFLC students may obtain an Associate of Arts degree through the joint DLIFLC-Monterey Peninsula College degree program.

The Defense Language Institute publishes two journals:

- *Applied Language Learning* (US ISSN 1041-6791) provides a forum for the exchange of ideas and information on instructional methods and techniques, curriculum and materials

development, assessment of needs within the profession, testing and evaluation, and implications and applications of research from related fields such as linguistics, education, communications, psychology, and the social sciences. It is published semiannually by the Defense Language Institute Foreign Language Center and Presidio of Monterey, and is available on the Internet and the Intranet of the Defense Language Institute Foreign Language Center. See <http://www.dliflc.edu>

- *Dialog on Language Instruction* is an occasional internal publication of the Defense Language Institute Foreign Language Center (DLIFLC) and part of its professional development program. Its primary function is to promote the exchange of professional information by increasing opportunities to share knowledge among DLIFLC faculty and staff and by encouraging professional communication within the worldwide Defense Foreign Language Program. See http://www.dliflc.edu/Academics/outside_ref_aj.html

Other DLIFLC initiatives include the *Emerging Languages Task Force* (ELTF). Its mission is to provide a rapid response in establishing new language programs in low-density languages in support of the Global War on Terrorism. This consists of establishing resident programs for language instruction, testing, familiarization training for deploying units, technology testing, as well as performing special translation projects.

The Defense Language Institute Foreign Language Center (DLIFLC) set up the ELTF (originally named the Operation Enduring Freedom Task Force, then later, the Global War on Terrorism Task Force) shortly after September 11, 2001 in response to Department of Defense needs for language capability in certain less-commonly taught languages. The ELTF first began teaching Persian-Afghan (Dari), Pashto, Georgian and Uzbek. Since then Kurdish has been added to the list of languages taught, and the list continues to grow. The ELTF continues to serve as DLIFLC's quick-response language team, which provides adequate solutions to current and emerging mission needs for the Global War on Terrorism.

University of Wisconsin-Madison Center for Southeast Asian Studies (CSEAS)

The Center for Southeast Asian Studies was formally established in 1973 as an interdisciplinary program of the College of Letters and Sciences. Formal scholarly interest in Southeast Asia, however, developed in the early 1950s when Wisconsin pioneered cooperation in economics with the University of Indonesia and Gajah Mada University. Subsequently, the program expanded considerably as scholars with research and teaching experience in the region joined the faculty.

In the early 1960s, the University established a combined East and Southeast Asia Program, from which the Center for Southeast Asian Studies emerged as a separate program in 1973. Since 1981, the Center has been recognized as a National Resource Center by the U.S. Department of Education and has received Title VI funding for program development and Foreign Language and Area Studies (FLAS) fellowships. In addition to federal support, the Center's program has benefited greatly from substantial grants from the Henry Luce Foundation (1988-93) and from a million dollar challenge grant from the Mellon Foundation.

In 1996, the Center participated in the establishment of the International Institute and currently collaborates with other area programs in global and comparative research and activities. At present, the Center operates a vigorous and well-developed program committed to the study of Southeast Asia that is recognized as one of the best worldwide.

At present the program consists of 25 core faculty members in 15 disciplines that include anthropology, business, communications (journalism), education, economics, forestry, geography, history, linguistics, literature, music and dance (performing arts), political science, public health, sociology, and urban and regional planning (see Faculty & Staff). Many other University faculty members have research and teaching interests in Southeast Asia and regularly work with students in a wide variety of fields, including environmental, development, and policy studies, natural resources, and law.

More than 120 undergraduate and graduate students specializing on Southeast Asia are enrolled in various departments and institutes throughout the University. The Center administers two major degree programs: BA in Asian Studies (Southeast Asian Concentration) and MA in Southeast Asian Studies (see Academic and Degree Programs).

Some sixty courses with significant Southeast Asia content are offered in various departments and colleges each academic year (see Southeast Asia Course Offerings). Regular instruction is offered through the Department of Languages and Cultures of Asia in three Southeast Asian languages: Filipino/Tagalog, Indonesian, and Thai. In addition, the Center is actively involved in the four advanced language training programs (for Filipino, Indonesian, Thai, and Vietnamese) held in Southeast Asia during the summers (see Language Programs) and has been chosen to host the Southeast Asian Studies Summer Institute (SEASSI) from 2000-2009. Faculty and student research are supported by extensive library holdings, with particular strengths on Indonesia, the Philippines, and Thailand, including a separate Southeast Asian Video Archive and other special collections (see Resources for Study and Research).

The Center offers two varieties of fellowship support for graduate students: Foreign Language and Area Studies awards (FLAS, funded by the Center's Title VI grant) and Center Fellowships (see Financial Aid). Considerable emphasis is placed on study abroad and field research in Southeast Asia. All graduate students are encouraged and assisted in their efforts to conduct field research in the region, and many are successful in the national competition for dissertation research grants. The University administers an undergraduate study abroad program in Southeast Asia, the College Year in Thailand program.

In 1998, the Center also inaugurated a graduate student study abroad program in collaboration with Thammasat University in Bangkok (see Overseas Programs). To encourage interaction with Southeast Asian scholars and to enrich the on-campus program, visiting faculty from the region are regularly invited. In recent years scholars from Indonesia, Thailand, Vietnam, Japan, and Australia have been in residence for varying periods of time.

Through language training and related area studies education, Southeast Asian studies at Wisconsin has contributed to the nation's wealth of foreign area expertise. Since 1980 more than sixty Ph.D. and over eighty M.A. degrees have been awarded to students conducting research on Southeast Asia. Employment opportunities for these graduates have varied, with roughly half of the Ph.D. degree holders having found academic teaching positions in the United States or abroad and many others securing careers in administration or research with private, government, or international organizations. The Center remains committed to its goal of assisting talented and deserving students in gaining the education needed for scholarly and professional careers in Southeast Asian studies.

Center Website The Center supports and maintains its own web page at the url:

www.wisc.edu/ctrseasia. This page provides relevant and up-to-date information on the Center and its programs and activities. The site features easy access to such Center-related links as the Friday Forum lecture series, information on faculty, Southeast Asia course offerings, the SEAiT project, language programs, and the Center's publications program.

Southeast Asian Summer Studies Institute (SEASSI)

(hosted by CSEAS and the University of Wisconsin-Madison 2000-2009)

SEASSI was established in 1983 to provide high quality intensive language instruction in at least eight of the major Southeast Asian languages at one location during the summer. It was and remains a unique program of area language training that combines institutional, federal, and foundation funding to bring faculty and students together in a single location every summer.

SEASSI is an integral part of a nationwide network of language teaching faculty from the institutions that are members of the SEASSI Consortium: fourteen North American universities committed to promoting Southeast Asian language and area studies. These universities pool their resources, including U.S. Department of Education Foreign Language and Area Studies (FLAS) fellowship funding, to support language instruction at SEASSI.

The members of the SEASSI Consortium are: Arizona State University, California State University-Long Beach, Cornell University, Northern Illinois University, Ohio University, University of British Columbia, University of California-Berkeley, University of California-Los Angeles, University of Hawaii-Manoa, University of Illinois at Urbana-Champaign, University of Michigan, University of Washington, University of Wisconsin-Madison, and Yale University.

Instruction

SEASSI offers an eight-week intensive language training program for undergraduate and graduate students and professionals. Instruction is offered in the following languages at the first, second and third year levels: Burmese, Hmong, Indonesian, Javanese, Khmer, Lao, Tagalog, Thai and Vietnamese. Scheduling of classes is contingent on sufficient enrollment, especially at the upper levels.

Each language course is equivalent to two semesters of study, with full academic year credit. Instruction is given in small individualized groups taught by a team consisting of a coordinator (usually a linguist specializing in Southeast Asian language pedagogy) and teachers who are native speakers of that language.

Heritage Language Instruction: Hmong, Khmer, Lao, Vietnamese, and Filipino

SEASSI offers specialized courses to Southeast Asian American students. Like all SEASSI courses, the heritage language courses are equivalent to two semesters of study, with full academic year credit.

Southeast Asian Area Studies

Every summer several Southeast Asian courses are offered. During the summer of 2004, SEASSI offered the following area studies courses to SEASSI students: Javanese gamelan (Music 361, 1 credit); Looking into Island Southeast Asia: Films as Windows to Understanding (Interdisc. 202, 3 credits); and Non-credit Lecture Series: Islam in Southeast Asia. SEASSI also organized an annual poetry night, a Southeast Asian music festival, a weekly film series, and two conferences: (1) Council of Teachers of Southeast Asian Languages (COTSEAL)'s annual meeting and (2) the SEASSI graduate student conference.

APPENDIX 8. RESUMES OF KEY PERSONNEL

DOUG COOPER, Project Director and Software Developer

PROF. ROBERT J. BICKNER, Senior Curriculum Consultant

FRANK SMITH, Lexical Resources Manager

PROF. THOMAS GETHING, Outside Reviewer

KITIMA INDRAMBARYA, Linguist Consultant

TITIMA SUTHIWAN, Linguist Consultant

PONGSKORN SAIPETCH, Software Consultant

LWIN MOE, IT Systems Manager

Doug Cooper — Academic Publications and Employment

Born: July 29, 1953

Email: doug@th.net

Research Interests

— Problems underlying software development for Southeast Asian languages, including text-to-speech, optical character recognition, phonetic approximation, parallel alignment of bilingual texts, automatic segmentation, text corpora, and design of metrics for performance / quality evaluation.

— Development of digital lexical resources for Southeast Asia, including comparative dictionaries for the Tai and Mon-Khmer language families, tools for working with epigraphic texts, and tools for Southeast Asian etymological research, with a particular focus on enabling access to Sanskrit, Dravidian, and Chinese sources.

Employment

2006 – 2009 Technical Director (40% appointment), *SEAlang Library* (Prof. Robert Bickner, Project Director, University of Wisconsin-Madison, lead institutional collaborator)

2003 – present Executive Director, *CRCL Inc.*

1995 – 2003 Head of the informal *Center for Research in Computational Linguistics*.

1989 – 1990 *Smith College*, Lecturer, Dept. of Computer Science,.

1984 – 1988 *University of California, Berkeley*, Lecturer, Dept. of Electrical Engineering and Computer Sciences.

Note: throughout the period 1979 – 1994 almost all of my time was devoted to writing university-level computer science textbooks.

Textbooks

— *Teaching Introductory Programming, 2nd Edition*, 1993 (504 pages)

— *Oh! Pascal!, 3rd Edition*, 1993 (lvi + 647 pages)

— *Oh! Pascal! for Turbo Pascal 6.0, 3rd Edition*, 1992 (lxi + 836 pages)

— *Oh My! Modula-2!*, 1990 (xxxii + 754 pages)

— *Condensed Pascal, with more Math and Science*, 1987 (xxviii + 612 pages)

— *Teaching Introductory Programming*, 1985 (x + 173 pages)

— *Oh! Pascal!, 2nd Edition*, 1985 (lv + 540 pages)

— *Standard Pascal User Reference Manual*, 1984 (iv + 183 pages)

— *Oh! Pascal!*, 1982, with Michael Clancy (xxx + 486 pages)

All texts published by W.W. Norton & Co., New York and London. Some texts are available in German, Chinese, and Spanish translation. *Oh! Pascal!* is available for blind students in Braille, prepared by the author. *Standard Pascal User Reference Manual* is available for blind students in Braille, prepared as above, and on audio tape.

Selected Papers, Technical Reports, and Presentations

- *New Tools for non-Roman L2s: extensive reading, intensive vocabulary acquisition, and basic research.* Invited plenary speaker, Interagency Language Roundtable, Washington DC, September 30, 2005
- *SEALANG Lab: 3/3/3 + Surge Capacity for Southeast Asian Languages.* ILR Showcase, National Cryptologic School, July 29, 2005
- *Beyond Borders: the SEAlang Library* National Library of Thailand Centennial, Bangkok, Thailand. July 6-8, 2005
- *SEAlang Library: Digital Texts and Tools for Southeast Asian Studies.* Asian Studies Information Access Roundtable, Association for Asian Studies, April 2005.
- *How to Read Less and Know More: Approximate Optical Character Recognition for Thai* (Association for Computing Machinery SIGIR '97, Philadelphia, USA)
- *45,656 Thai Names: Statistics and Implications of passlist.96* (1997 International Conference on Computer Processing of Oriental Languages, Hong Kong)
- *Characteristics of Ambiguous (Partition(s)in) Thai Text* (11th Pacific Asia Conference on Language, Information, and Computation, Seoul, Korea)
- *The Thai Internet Education Project* (accepted Information Technology in Selected Southeast Asian Countries).
- *Information Management in Central Southeast Asia: What's not Happening (and Why)* (accepted Pacific Asia Conference on Information Systems 1997, Brisbane, Australia)
- *Can Information Technology Survive in Thailand?* (accepted 1997 Information Resources Management Association International Conference, Vancouver BC Canada)
- *What can Information Technology Offer Khmer Studies?* (International Conference on Khmer Studies, Phnom Penh University, 1996)
- *The Brain Game — Policy Issues in Building Thailand's Research Infrastructure* (Sixth International Conference on Advanced Science and Technology Exchange with Thailand, Bangkok, 1996)
- *Compared to What? Measuring the Accuracy of Thai Text Segmentation Algorithms* (Paper 9, Center for Research in Computational Linguistics)
- *Building Southeast Asia's Research Infrastructure: Cooperative, Public-Domain Research on the World-Wide Web* (15th Southeast Asian Regional Computing Conference '96, Bangkok)
- *Who's Bringing the Information to the Information Highway? University Research Policy and the Internet* (International Conference on Reengineering the University, March, 1996, Chulalongkorn University, Bangkok)
- *How Do Thais Tell Letters Apart? A Study of the Secondary Characteristics of Thai Letters* (Pan-Asiatic Linguistics 1996 — 4th International Symposium on Language and Linguistics, Mahidol University)
- *Sorting by Sound — Arbitrary Lexical Ordering for Transcribed Thai Text* (10th Pacific-Asia Conference on Language, Information, and Computation, City U. of Hong Kong 1995)
- *Font Design for Thai/English Typesetting* (Symposium on Natural Language Processing '95, Kasetsart University)
- *Fuzzy Letters and Thai Optical Character Recognition* (SNLP '95, Kasetsart University)

— *How to Sort Thai Without Rewriting sort* (Paper 6, Center for Research in Computational Linguistics)

Miscellaneous

Wrote *SEAlang Library* proposal; funded 2006-2009 (\$432,000) by USED *TICFIA* (Technical Innovation and Cooperation for Foreign Information Access) program.

List owner *SEALANG-L* (Southeast Asian Languages) and *NLPASIA* (Natural Language Processing in Asia) international mailing lists, maintained by NECTEC (National Electronics and Computer Technology Center), Thailand. Initiator of the *SEALEX* (Southeast Asian Lexicography) SIG.

Consultant: Microsoft / Monotype, 1998, design of Thai portion of standard Tahoma font.

Consultant: Sun Microsystems, Digital Equipment Corporation, City of Berkeley, California

Consulting Editor/Reviewer: W.W. Norton & Co., John Wiley & Sons, Prentice-Hall Publishers, Addison-Wesley Publishing.

Grant: 1987 proposal for *Research in Undergraduate Education* awarded \$1.2 million by Digital Equipment Corporation.

Member: International Standards Organization/American National Standards Institute Committee (X3J9) on the programming language Pascal.

Invited speaker, ACM/SIGSCE 1992: *Issues in Choosing a Language for the Introductory Programming Course*

Robert J. Bickner

Dept. of Languages and Cultures of Asia
University of Wisconsin
Madison, WI 53706
(608) 262-3915, 265-3538 (fax)

521 Grace Court
Verona, WI 53711
(608) 845-5376

Employment

- 1994 - present Professor (Thai Language and Literature)
Department of Languages and Cultures of Asia
University of Wisconsin – Madison
Project Director, SEAlang Library (2006 – 2009)
- 1989 - 1994 Associate Professor (Thai Language and Literature)
Department of Languages and Cultures of Asia
University of Wisconsin - Madison
- 1983 - 1989 Assistant Professor (Thai Language and Literature)
Department of Languages and Cultures of Asia
University of Wisconsin - Madison
- 1981 - 1983 Visiting Assistant Professor (Thai Lang. and Lit.)
Department of Languages and Cultures of Asia
University of Wisconsin - Madison
- 1970 - 1974 U.S. Peace Corps Volunteer
Instructor of EFL and Teaching Methodology
Chombung Teacher College, Ratchburi, Thailand

Education

- PhD 1981 University of Michigan, Major field: Linguistics
- MA 1977 University of Michigan, Major field: Asian Studies
- BA 1969 Catholic University of America, Major field: English Literature

Language Competence

Fluent in Thai, some competence in, or abstract knowledge of: Cebuano, French, Hmong, Khmer, Lao, Latin, Sanskrit, Vietnamese.

Committee & Service Work

- 1982-present College Year in Thailand Program (CYIT), Director
- 2000-present Southeast Asian Studies Summer Institute (SEASSI), Language Director
- 1993- 2000 Chair, Department of Languages and Cultures of Asia, UW-Madison

Books and Monographs

- 1991 *An Introduction to the Thai Poem "Lilit Phra Law" (The Story of King Law)*, Monograph Series on Southeast Asia, Center for Southeast Asian Studies, Northern Illinois University, 254 pages.

Electronic Media Materials

- 1996 *Computer Assisted Instruction for Thai Reading*, a 32 lesson Macintosh HyperCard Set

Articles

- 2002 "Reflections on a Literary Dispute Between Jit Phumisak and Phra Worawetphisit," *Proceedings of the 10th Annual Southeast Asian Linguistics Conference of the Southeast Asian Linguistic Society 2000*. Arizona State University Program for Southeast Asian Studies, Monograph Series Press, 87-96.
- 1992 "Some Textual Evidence on the Tai Sounds *ai and *aʰ," *Papers on Tai Languages, Linguistics, and Literatures In Honor of William J. Gedney on his 77th Birthday*, Carol J. Compton and John F. Hartmann, eds., Occasional Paper No. 16, Center for Southeast Asian Studies Northern Illinois University, 223-230.

- "The Nature of 'Standard' Thai," (with Thomas John Hudak), *Journal of South Asian Literature*, 25.1, 163-175.
- "The Problems of Southeast Asian Language Instruction: The Case of Thai," in *Southeast Asian Studies in the Balance: Reflections from America*, Charles Hirschman, et al., eds. Association for Asian Studies, 93-123.
- 1989 "Directional Modification in Thai Fiction: The Use of 'Come' and 'Go' in Text Building," David Bradley ed., *Papers in South - East Asian Linguistics* No. 11: *Southeast Asian Syntax*, 15-79. *Pacific Linguistics*, A-77, The Australian National University.
- "Cultural Variation in Reflective Writing," (with Patcharin Peyasantiwong) *Writing Across Languages and Cultures: Issues in Contrastive Rhetoric, Written Communication Annual Vol. II*, Alan C. Purves, ed., Sage Publications, 160-74.
- 1986 "Thai Tones and English Loanwords: A Proposed Explanation," *Papers from a Conference on Thai Studies in Honor of William J. Gedney*, Robert J. Bickner, et al., eds. Michigan Papers on South and Southeast Asia. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan, 19-39.
- "Changing Perspectives on Language and the Poetic Arts in Thailand," *In Celebration of King Ram Khamhaeng of Sukhothai: Seven Hundred Years of the Development of the Thai Language*, J. Hartmann, ed. CROSSROADS, Center for Southeast Asian Studies, Northern Illinois University, 3.1, 104-117.

Books Edited

- 1988 *Selected Writings in Tai Linguistics by William J. Gedney*, R. Bickner et al, eds., Michigan Papers on South and Southeast Asian Studies. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- 1986 *Papers from a Conference on Thai Studies in Honor of William J. Gedney*, R. Bickner et al, eds., Michigan Papers on South and Southeast Asian Studies. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Presentations & Consulting

- 2004 "Minority Adaptation in Thailand," Wisconsin Dept. of Instruction pre-departure seminar for 18 academic personnel traveling to Thailand to study refugee populations.
- -- Evaluator for the 2004 *Tuttle Asian Language Publication Grant*, Tuttle Publishing.
- 2003 "Doing Business in Thailand," Wisconsin Dept. of Commerce, International Division.
- 2002 Translation and Interpretation services for Lutheran Social Services and Dane County Circuit Court, case No. 02 TP 42.
- 2000 "Reflections on a Literary Dispute Between Jit Phumisak and Phra Worawetphisit," the 10Th Annual Southeast Asian Linguistics Conference of the Southeast Asian Linguistic Society, Madison, May 4th - 7th.
- 1997 "Culture and Religions of Southeast Asia as They Affect Contemporary Issues," Wisconsin Rural Leadership Program Conference, LaCrosse WI.
- 1991 "The Nature and Observance of Holidays in the Buddhist Cultures of Southeast Asia," Invited presentation for Madison Metropolitan School District, teacher in - service training program, November 20th. The lecture was subsequently broadcast several times over local cable television.
- "Opening Doors to Thailand," for a session entitled *Diversifying Destinations: Southeast Asia*, 44th International Conference on Educational Exchange, Denver, Nov. 6-8.
- "Stereotypical Views: East and West and Cross - cultural Adaptation," UW - Madison Center for Southeast Asian Studies, Friday Forum, April 19.
- "The Problems of Southeast Asian Language Instruction: The Case of Thai," panel on Southeast Asian language instruction, Association for Asian Studies, New Orleans, April.

- "Issues and Goals for the Development of Instructional Programs in Southeast Asian Languages," one of four invited presenters for a round table discussion organized by the Consortium of Teachers of Southeast Asian Languages (COTSEAL), Association for Asian Studies, April.
- 1990 "The Problems of Southeast Asian Language Instruction: The Case of Thai," invited presentation for a conference entitled "Southeast Asian Studies and the Social and Human Sciences," cosponsored by the Association for Asian Studies and the Joint Social Science Research Council/American Council of Learned Societies Committee on Southeast Asia, with funding from the Johnson, Ford and Luce Foundations, at the Wingspread Conference Center, Racine, WI..
- 1989 "Language and Culture in Asia," conference on "Linguistics and Language Learning and Teaching" at Payap University, Chiang Mai Thailand.
- "Student and Faculty Exchange Programs with the United States," Khon Kaen University, Khon Kaen, Thailand..
- "Translation of the Thai Literary Classic: Lilit Phra Law", Mahasarakham Teacher College, Mahasarakham, Thailand, March 7th, 1989.
- "Subtlety in Communication; English vs. Thai," for a conference on "Teaching English in Its Cultural Context," Prince of Songkhlaa University, Pattani, Thailand, March 22nd, 1989
- 1988 "Some Observations on a Thai Village Ritual," Slide/Lecture presentation, also given in 1981, '82, '83, '85, and '87 for various audiences and classes.
- 1986 "Plot and Structure: The Catalog as a Coherence Device in "Lilit Phra Law," Council on Thai Studies, Madison.
- "The Traditional Poetry of Thailand," Asian Poetry Evenings Series, an invited presentation for the Asia Society, New York.
- "Artistic Expressions of the Thai," Peoples and Cultures of Southeast Asia Seminar, Residential Weekend Seminar Series, UW - Extension, Madison.
- 1985 "Some Textual Evidence on the Age of the Thai Vowel Distinction *ai vs. *aʔ," 18th International Conference on Sino - Tibetan Languages and Linguistics, Bangkok.
- 1984 "Linguistic Analysis and Thai Literary Studies: the Need for a New Beginning," invited lecture, the International Symposium on Language and Linguistics, Chiang Mai University, Thailand.
- 1983 "Literature and Literacy in Thailand: a Discussion of Some Significant Misunderstandings," Council on Thai Studies, DeKalb.
- "Literary Studies and Historical Reconstruction: The Case of Thai," 17th International Conference on Sino - Tibetan Languages and Linguistics, Seattle.
- "The Literature of the Chakri Dynasty of Thailand," discussant for the Rattanakosin (Bangkok) Bicentennial Conference, DeKalb.
- "Thai Literature and Historical Linguistics," invited presentation, Southeast Asian Program and Department of Anthropology, University of Chicago.
- "Adaptation of Indic Verse Forms in Thai," Conference on South Asia, UW - Madison.
- Panelist, Adaptation Problems of Hmong Refugees, "The State of Wisconsin" program, WHA television.
- "Refugees and Adjustment: Background Information and Appropriate Strategies for Local Service Providers," Regional Conference for Returned Peace Corps Volunteers, Madison.
- 1981 "Indochinese Refugees in America: Linguistic and Cultural Problems," Teachers of English to Speakers of Other Languages (TESOL), Midwest Regional Meeting, Urbana.
- 1980 "The Position of *chan* Poetry in the Literary Heritage of Thai," Conference on Thai Studies in Honor of William J. Gedney, Ann Arbor.

Curriculum Vitae for Frank Smith

PO Box 260098
Madison, WI 53726-0098
USA

Tel. (Thailand, Sept-May) +66 9-014-8105
(US, June-Aug) (608) 263-1755
E-mail: fsmith@wisc.edu

Professional Experience

Teacher Trainer

Center for Khmer Studies

Phnom Penh, Cambodia

October, 2004 and October, 2005

Designed and conducted a week-long workshop for 16 Cambodian teachers of Khmer language to foreigners. Instructed workshop participants, using hands-on, participatory methods, in classroom planning, teaching techniques and materials design. Conducted entirely in Khmer.

Teacher Trainer

“Teaching Southeast Asian Languages: Performance-based Instruction”

University of California-Los Angeles

Los Angeles, CA

January, 2003

Co-wrote grant proposal, co-designed overall workshop and presented four individual workshops in a three-day teacher training workshop run by the Council for Teachers of Southeast Asian Languages.

Heritage Language Program Facilitator

Southeast Asian Studies Summer Institute

University of Wisconsin-Madison

August, 2001 – Present

Coordinate curriculum design and instruction for Khmer Heritage Language Program at intensive summer language institute. Consult with and provide technical assistance to coordinators of other Heritage languages offered (Hmong, Vietnamese, Lao) to aid in curriculum design and unified vision/scope of Program. Design and maintain Heritage Program website: <http://www.seassi.wisc.edu/heritage/index.html> Recruit university students throughout the U.S. for Program via e-mail and occasional campus presentations.

Khmer Language Coordinator

Southeast Asian Studies Summer Institute

Various Universities

1990, 1993, 1996-7, 2000-Present

Oversee all aspects of intensive summer Khmer language program at Cornell University (1990), University of Washington-Seattle (1993), Arizona State University (1996-7) and University of Wisconsin-Madison (2000-4). Hire, train and supervise native-speaking Khmer instructors. Design curriculum and teaching materials for all levels of instruction. Teach Khmer in performance-based, communicative language classrooms at all levels. Counsel learners. Design, author and maintain extensive website of Khmer instructional materials (*URL and password available on request*).

ESL Curriculum Designer, Healthy Boston Project**City of Boston Mayor's Office****Boston, MA****June, 1994 – August, 1994**

Designed and authored a 10 episode cable television program for adult ESL learners, entitled *Speak Easy*, with a topical focus on health issues. Each episode featured scripted dialogue for non-native speaking actors followed by a scripted on-screen teacher presentation of grammatical material and vocabulary, presented in context via clips from the dialogues. *VCD of sample episode available on request.*

ESL Instructor**Harborside Community Center****Boston, MA****May, 1991 – May, 1994**

Designed all curriculum and materials for and taught high beginning/low intermediate adult ESL class at an urban community school. Also (1991-1992) designed and taught a bilingual ESL class for Khmer adults. Contributed to design of program-wide intake assessment instrument.

Education***MLS (Master's in Library Science)*****Simmons College****Boston, MA****January 1994 – December, 1994**

Specialized coursework in multimedia (including website) design, online information and database management. **GPA: 3.85**

BA in Sociology (honors)**University of Wisconsin-Madison****August, 1985 – August, 1988**

Significant coursework in Anthropology as well as Sociology. Completed 168-page Senior Honors Thesis on Cambodian peasant refugee lifeways in U.S., based on ethnographic research funded by a Knapp-Brittingham Honors Research Grant. **GPA: 3.75**

Publications

Beginning Lao Study Guide. Madison, WI: Southeast Asian Studies Summer Institute, 2005.

Than Samay: A Lao Heritage Language Textbook for University Level. Madison, WI: Southeast Asian Studies Summer Institute, 2005.

Khmer Language Training Guidelines for Wisconsin DPI Instructors. Madison, WI: Wisconsin Department of Public Instruction Document, Feb., 2003.

“Review of ‘Soul Survivors: Stories of Women and Children in Cambodia,’ by Carol Wagner, et. al.,” in *Journal of Asian Studies*, Vol. 62, No. 1 (Feb., 2003).

Heritage Khmer: An Introduction Through Literacy. Madison, WI: Southeast Asian Studies Summer Institute, 2002. Includes CD-ROM, Audio CD and workbook.

Sophat: Novel and Study Guide. Madison, WI: Southeast Asian Studies Summer Institute, 2002.

Intensive Khmer Course: Beginning. Madison, WI: Southeast Asian Studies Summer Institute, 2001. Includes CD-ROM and Audio CD.

Intermediate Khmer Reading Using Roadside Signs (website):
<http://carla.acad.umn.edu/lctl/khmer/signs.html> 2002.

“Cultural Consumption: Cambodian Peasant Refugees and Television in the ‘First World,’” in May Ebihara, Carol Mortland and Judy Ledgerwood (eds.), *Cambodian Culture Since 1975: Homeland and Exile*. Ithaca, NY: Cornell University Press, pp. 141-160. 1994.

Interpretive Accounts of the Khmer Rouge Years: Personal Experience in Cambodian Peasant World View. Madison, WI: Center for Southeast Asian Studies, University of Wisconsin-Madison. Wisconsin Papers on Southeast Asia, Occasional Paper No. 18. 1989.

Conference Presentations

Telling the Customer What to Do: A Thai Sex-Worker Folk Chant. Paper presented at the **Association of Asian Studies Annual Meeting**, Chicago, IL, March, 2005.

A Multimedia Content Approach to Khmer Heritage Language Instruction. Poster Session presented at the **2nd Heritage Languages in America Conference**, Tyson’s Corner, VA, October 18-20, 2002.

Teacher Training by Example: a DVD Classroom Demonstration Project. Multimedia presentation at the **12th Annual Conference of the Council of Teachers of Southeast Asian Languages**, University of Wisconsin-Madison, Madison, WI, July, 2002.

Sight Words and Captured Birds: Multi-Methodologies in the Teaching of Khmer Literacy. Paper presented at the **Association of Asian Studies Annual Meeting**, Washington, DC, April, 2002.

Comic Worlds: Artwork-mediated Involvement, Everyday Storytelling and Pure Spoken Practice in the Intermediate Khmer Classroom. Paper presented at the **9th Annual Conference of the Council of Teachers of Southeast Asian Languages**, University of Oregon, Eugene, OR, July, 1999.

Languages

English (native); Khmer (distinguished spoken proficiency and literacy); Thai and Lao (advanced spoken proficiency and literacy); French (basic literacy only).

Thomas W. Gething

Affiliate Professor and Acting Associate Dean
University of Washington, Thomson Hall 303
Seattle Washington 98195

getting@u.washington.edu

B.A.1961, English Language, University of Michigan (with distinction)

M.A.1963, Linguistics, University of Michigan

Ph.D.1966, Linguistics, University of Michigan

Attended Indiana University 1964 and Yale University 1964-65

Awards and Grants

NDFL fellow (Title VI) 1962-1963 and 1963-1964

ACLS Summer Study Grant recipient 1964

Foreign Area Training Program (ACLS-SSRC) fellow 1964-1966

U. S. Office of Education research grant recipient 1976-1977

Award for the Enhancement of Graduate Studies and Student Life 1980

Fulbright Program research grantee 1982

U.S. Office of Education materials project director 1984

U.S. Department of Education Group Projects Abroad grants 1992-1995, 1999-2007

Henry Luce Foundation grant 1996

Academic Appointments

1962-1964	English Language Institute and Peace Corps Thailand I and II training programs, University of Michigan
1966-1967	Department of Linguistics, University of Michigan (Assistant Professor)
1967	Linguistic Society of America, Summer Linguistic Institute, Ann Arbor
1967-1970	Department of Asian and Pacific Languages, University of Hawaii (Assistant Professor)
1970-1971	Department of Linguistics, Ohio University (Associate Professor)
1971-1995	Department of Indo-Pacific Languages, University of Hawaii at Manoa (Associate Professor, Professor)
1973	Linguistic Society of America, Summer Linguistic Institute, Ann Arbor
1977	Linguistic Society of America, Summer Linguistic Institute, Honolulu
1995-present	Department of Asian Languages & Literature, University of Washington (Affiliate Professor)
2000-2002	National Foreign Language Center, University of Maryland (Senior Affiliate)

Consultantships (recent)

University of Wisconsin (Madison); program evaluation (1990)

Consortium for International Development, Tempe, Arizona; teacher education curriculum project in Lao People's Democratic Republic (1993)

U.S. National Security Education Program; evaluator (1994 and 1995)

U.S. National Endowment for the Humanities; grant reviewer (1994)

U.S. Fulbright Program Scholar-in Residence reviewer (1993-1996, 2000)

Georgetown University Title VI impact assessment [for Southeast Asian languages] (1995)

University of Washington (Seattle); program evaluation (1995-1996)

U.S. Department of Education; IEGPS grants panel member (1997-present)

Institute for International Education; grants panel member (1998-2002, 2005)

U.S. Department of Education; OERI consultant on "College Course Map" (2001)

Cornell University; program evaluation (2004-2005)

Dissertation Committees (recent: University of Hawaii)

- 1995 Maneewan Pewnim, "Changes in Leisure Activities in a Thai Provincial City"
Christine Des Jarlais, "Faculty Morale in an American Public Research University"
- 1996 Titima Suthiwan, "Malay Lexical Elements in Thai"
- 1997 Varisa Osatananda, "A Study of Prosodic Systems in Vientiane Lao"
- 1999 Umayah Binti Haji Umar, "A Classification of Thai-Kedah Dialects Using Phonological Characteristics" [external examiner for Universiti Malaya]
- 2000 Rainer Stasiewski, "Phonological Interference between Tone Languages: Chinese and Thai"

Publications (selected)

- 1975 (with R. B. Jones) Languages of Southeast Asia. Material development needs in the uncommonly-taught languages. Arlington, Virginia: Center for Applied Linguistics, pp. 76-9.
- 1976a (co-editor with Harris and Kullavanijaya) Tai linguistics in honor of Fang Kuei Li, Bangkok: Chulalongkorn University Press, 254 pp.
- 1976b. Notes on Lao personal pronouns. Tai linguistics in honor of Fang Kuei Li, Gething, Harris, and Kullavanijaya, editors. Bangkok: Chulalongkorn University Press, pp. 103-12.
- 1977 (with P.T. Bilmes) Teacher's manual for Thai basic reader. Honolulu: Department of Indo-Pacific Languages, University of Hawaii, 78 pp.
1978. (with P.T. Bilmes) Thai basic reader. Honolulu: Southeast Asian Studies Program. University of Hawaii, 295 pp.
- 1985a. (with P.T. Bilmes) Thai for public health professionals. Honolulu: Southeast Asian Studies Program, University of Hawaii, 252 pp.
- 1985b. The Thai language as a map of Thai culture. Papers from a Conference on Thai Studies, Bickner, Hudak, and Peyasantiwong, editors. Ann Arbor: Center for South and Southeast Asian Studies, #25.
- 1986a. Selective development of the Thai lexicon, Crossroads, v. 3, no. 1, pp. 118-122.
- 1986b. Thai semantic structure and Thai culture, Indo-Pacifica, nos. 2 and 3, Honolulu: University of Hawaii, pp. 77-89.
- 1986c. Review of Poetics of the Ramakian by Bofman. Journal of the Association for Asian Studies, v. 45, no. 5, pp 1116-7.
1989. Li Fang-Kuei entry in International dictionary of anthropologists, Winters, editor. Garland, New York.
- 1992a. (with P.T. Bilmes) Thai basic reader (Revised). Honolulu: Center for Southeast Asian Studies, University of Hawaii at Manoa, 336 pp.
- 1992b. (with P.T. Bilmes) Teacher's manual for Thai basic reader (Revised). Honolulu: Center for Southeast Asian Studies, University of Hawaii at Manoa, 95 pp.
1995. Review of Tai languages, linguistics, and literature by Compton & Hartmann. Journal of the Association for Asian Studies, v. 54, no. 1, pp. 253-4.
2000. Lao entry in Facts about the world's languages; An encyclopedia of the world's major languages, past and present, Garry and Rubino, editors. H.W. Wilson Press, New York and Dublin, pp. 409-11.

Titima Suthiwan
 Assistant Professor
 Deputy Director and Convenor for Thai Language Program
 Centre for Language Studies
 Faculty of Arts and Social Sciences
 National University of Singapore
 AS4/0105, 9 Arts Link, Singapore 117570
 Tel : (65) 6516 3706
 Fax : (65) 6777 7736
 email : clsts@nus.edu.sg

Education

- 1997 Ph.D. in Linguistics, University of Hawai'i at Manoa, U.S.A.
 1982 B.A. (Honors) in Thai Language and Literature
 Chulalongkorn University, Thailand

Papers and Publications

- (2005) **Song Kham** (*Duet*). Collection of poems in Thai, English, and French. Bangkok : Kaew Kham Publishing.
- 2005 **Loanwords in Thai : Time and Number**. Paper presented at the fifth workshop on Loanwords Typology. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- 2005 **Khian Thai: Thai Writing**. Singapore: Ridge Books, Singapore University Press.
- 2004 **Loanwords in Thai: Semantic change due to the introduction of Loanwords**. Paper presented at the third Workshop on Loanwords Typology. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- 2003 **Loanwords in Thai**. Paper presented at the Workshop on Loanwords Typology. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- 2002 **Kaan dern thang khawng jacob** (a translation of *Jacob's Journey* , written by Noah BenShea) Bangkok:Kobfai Publishing Project
- 2001 **Jacob khon tham khanom pang** (a translation of *Jacob the Baker*, written by Noah BenShea) Bangkok: Kobfai Publishing Project.
- A Linguistic Analysis of Malay Elements in Inao**. Paper presented at the Fifth International Symposium on Malay/Indonesian Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- 1999 **The Difference in Thai language Acquisition among Malay and Chinese Speaking Students at NUS**. Paper presented at COTSEAL conference,

- Southeast Asian Studies Summer Institute (SEASSI)99,
University of Oregon, U.S.A.
- 1997 **Malay Lexical Elements in Thai.** Ph.D. dissertation in
Linguistics, University of Hawai'i, U.S.A.
- 1996 **Thailand : Land of Contrasts.** A high school textbook.
Written jointly with Uri Tadmor. Honolulu: Center for
Southeast Asian Studies, University of Hawai'i.
- 1995 **An Analysis of Malay and Javanese Etyma in a Thai
Version of the Panji Cycle.** Working Papers in Linguistics , vol.25.
Honolulu : University of Hawai'i.
- 1994 Second edition of 1983
- 1992 **Malay Loanwords in Thai.** Proceedings of the Third
International Symposium on language and Linguistics, vol.3:
1358-1366. Bangkok: Chulalongkorn University
- 1992 **Kata Pinjaman Melayu dalam Bahasa Thai .** (*Malay
Loanwords in Thai*). Dewan Bahasa, vol. 36, 8:695-700
- 1989 **Linguistic Interference: The Case of Thai and
Khmer.** Written jointly with Uri Tadmor; presented to the
annual meeting of the Council on Thai Studies, Madison,
Wisconsin, U.S.A.
- 1987 **Maa Phro Khittheung.** (*Longing*). Collection of poems in
Thai. Bangkok : Kaew Kham Publishing.
- 1985 **Arai Kan Nak Kan Naa.** (*What's going on here*). A
children's book. Bangkok: Kaew Kham Publishing.
- 1983 **Yaak Ja Phiang Phuuk Phunn** (*Just attached*).
Collection of poems in Thai. Bangkok : Kaew Kham Publishing.
(Reprinted sixth time; No.1 annual bestseller)

Work Experience

- 2005- Deputy Director of Centre for Language Studies (CLS), Faculty
of Arts and Social Sciences (FASS), National University of
Singapore (NUS)
- 2003- Examiner Responsible for Language Thai A2, International
Baccalaureate Organization, Geneva, Switzerland.
- 2002- External Examiner in Thai, University of Hong Kong
- 2001-pres Assistant professor in and Convenor for the Thai language
program, CLS, FASS, NUS
- 1999 Director, Thai Study Abroad Program, Southeast Asian
Studies Programme (SEASP), FASS, NUS
- 1998 Assistant Professor, Southeast Asian Studies
Programme(SEASP), and Founder and Coordinator of the Thai
language program, SEASP, FASS, NUS

- Coordinator, Khmer, Lao, and Thai Language Programs,
Southeast Asian Studies Summer Institute (SEASSI),
University of Oregon, U.S.A.
- 1997 Thai language instructor, SEASSI 97, Arizona State
University, U.S.A.
- Coordinator, Thai language program, Department of
Hawaiian and Indo-Pacific Languages and Literatures
(HIPLL), University of Hawai'i at Manoa, U.S.A.
- 1993 Thai language instructor, SEASSI 93, University of
Washington, U.S.A.
- 1990-97 Lecturer, Thai language program, HIPLL, University of
Hawai'i at Manoa, U.S.A.
- 1989 Thai language instructor, SEASSI 89, University of Hawai'i
at Manoa, U.S.A.
- 1987 - 90 Graduate and teaching assistant, Department of Linguistics,
University of Hawai'i at Manoa, U.S.A.
- 1987 - 88 English language instructor, Bangkok University, Thailand
- 1985 English language instructor, Assumption College, Thailand
- 1983 - pres Freelance writer, translator, and interpreter

Scholarships and awards

- 1990 Henry Luce Fellowship for Southeast Asian Studies
- 1989 Henry Luce Travel Grant
Henry Luce Grant for SEASSI 1989
- 1983- 97 Pacific Area Studies Scholarship, University of Hawai'i at
Manoa
French poetry writing first prize award for intermediate level, Department
of European Languages, University of Hawai'i at Manoa

Current research

- Indic influence in Southeast Asian languages
Chinese influence in Southeast Asian languages
Women in Southeast Asian literature
Thai language acquisition

KITIMA INDRAMBARA, Ph.D**Assistant Professor**

Department of Linguistics,
 Faculty of Humanities, Kasetsart University
 50 Phaholyothin Road, Bangkok, THAILAND
 Email: fhumkmi@ku.ac.th

Education

- Dec 1994 Ph.D Linguistics, University of Hawai'i at Manoa.
- May 1989 M.A. Linguistics, University of Wisconsin at Madison.
- May 1989 Certificate in Teaching English as a Second Language, Department of English, University of Wisconsin at Madison.
- May 1987 B.A. English (Honors), Faculty of Arts, Chulalongkorn University, Bangkok.

Fellowships

- 1999 Thailand-Australia Science & Engineering Assistance Project (TASEAP) Fellowship for Computer Assisted Language Learning (CALL) training at Monash University, Melbourne, Australia
- 1991 – 1994 East-West Center Scholarship, Program on Cultural Studies, East-West Center, Honolulu, Hawai'i.
- 1990 American Association of University Women Scholarship, Honolulu, Hawai'i.

Professional Development

- 2004 – present Associate Dean of Information Technology, Faculty of Humanities, Kasetsart University
- 2000 – 2004 Chairperson, Humanities Computer and Network System Management Committee, Faculty of Humanities, Kasetsart University
- 2000 – 2004 Guest lecturer for an undergraduate program in English for at the Department of Foreign Languages, Faculty of Humanities, Kasetsart University
- 2000 – present Assistant Professor of the Department of Linguistics, Faculty of Humanities, Kasetsart University
- 1998 – 2002 Guest lecturer for a graduate program in English for Business Administration at the Department of Languages and Social Sciences, Faculty of Industrial Education, King Mongkut Institute of Technology, North Metropolitan Campus.
- 1997 – 2000 Chairperson, Self-Access Language Learning Centre Committee, Department of Foreign Languages, Kasetsart University
- 1999 – 2000 CALL Trainer in the topic of 'Software Selection' as part of Thailand-Australia Science & Engineering Assistance Project (TASEAP)
- 1999 Kasetsart University English language project liaison in the Thailand-Australia Science & Engineering Assistance Project (TASEAP)
- 1997 – 2000 Coordinator for the Technical English language course (Engineering), Department of Foreign Languages, Kasetsart University
- 1995 – 2000 Coordinator for the Foundation English I course, Department of Foreign Languages, Kasetsart University
- 1995 – 2000 Guest lecturer for a graduate program in Applied Linguistics, Department of Linguistics, Faculty of Humanities, Kasetsart University
- 1995 – 2000 Lecturer at the Department of Foreign Languages, Faculty of Humanities, Kasetsart University

Papers/Publications/Research

- 2005 “*Finding Subjects in Thai*” in **Manutsayasaat Wichakaan**, Faculty of Humanities, Kasetsart University.
- 2004 “*The Satisfaction of Users Towards the Faculty of Humanities Computer Room*”, Head of a joint research, supported by Faculty of Humanities, Kasetsart University.
- 2003 “*The Development of Computer-Assisted Interactive Program in Teaching English Through Songs for KU Students Taking Foundation English Courses*”, In **Kasetsart Journal: Social Sciences**. Vol.24, No.1
- 2003 “*The Computer-Assisted Interactive Program: Learn English Through Songs*”, In **Proceedings of 41th Kasetsart University Annual Conference** between 3 and 7 February 2003.
- 2003 “*The Satisfaction of Classroom Users Towards Classroom Buildings Organized by the Faculty of Humanities and Related Problems*”, Head of a joint research, supported by Faculty of Humanities, Kasetsart University.
- 2003 “*Feasibility Study for International Doctoral Program in Applied Linguistics, a joint research*”, supported by Faculty of Humanities, Kasetsart University.
- 2001 – 2003 “*An Institutional Research on Humanities Curriculum*”, a joint research, supported by Kasetsart University Research and Development Institute
- 2001 **English Structure II**. Kasetsart University Press.
- 2001 “*Kaan thàp sàp l Efl/ kaan lỳ̀p̄k si ð ang*” (*English loanwords and its pronunciation*) In **Manutsayasaat sahawithayakaan heng chiiwit** (in celebration of 20th anniversary of Faculty of Humanities), Kasetsart University
- 2000 “*On the Non-verbal Predicate in Thai.*” In **Grammatical Analysis: Morphology, Syntax and Semantics in Honor of Stanley Starosta**. Oceanic Linguistics Special Publication No.29. Videia P. De Guzman and Byron W. Bender. University of Hawai`i Press. 186-208
- 1997 “*The Status of the Word Hay in Thai;*” In **Pacific Linguistics A-90**: Ed. by M. Clark In Papers in Southeast Asian Linguistics No.16, 81-120.
- 1997 Sak-Humphry, Indrambarya, and Starosta, “*Flying In and Out in Khmer and Thai.*” **Southeast Asian Linguistic Studies in Honour of Vichin Panupong**. Ed. by Arthur S. Abramson. Bangkok: Chulalongkorn University Press. p.209-229.
- 1996 “*On Impersonal Verbs in Thai*” In **the Proceedings of the Fourth International Symposium on Language and Linguistics** Vol. I, Bangkok: Mahidol University Press. p.505-521.
- 1996 “*The Status of Auxiliary Verbs in Thai.*” In **the Proceedings of Southeast Asian Linguistic Society IV (SEALS IV)**, Bangkok: Ramkhamhaeng University.

Multimedia work/research

- 2004 *The Development of an E-courseware entitled ‘Word Factory Series: Phase I’*, supported by Faculty of Humanities, Kasetsart University
- 2000 – 2003 *The Development of Computer-assisted Interactive Program in Teaching English Through Songs for KU students Taking Foundation English Courses*, supported Kasetsart University Research and Development Institute
- 1996 – 1999 CAI Program on the Vocabulary Development of Engineering students at Kasetsart University, supported by Kasetsart University Research and Development Institute
- 1998 – 1999 Script writer and lesson designer for English I CAI program, the cooperative project between the Ministry of University Affairs and Kasetsart University.

Areas of Interest

Thai Linguistics, Syntax, Applied Linguistics, CALL, CAI

Dr. Pongsorn Saipetch, Ph.D.



Education

California Institute of Technology, Pasadena, California, U.S.A.
Bachelor of Science in Physics (with Honors), June 1991

University of California at Los Angeles, California, U.S.A.
Doctor of Philosophy in Biomedical Physics, June 1995

Management and Technical Experience

Managing Director, Atrium Technology Co., Ltd. (1995-)

High-performance Java development.

Secure Internet database systems.

Custom development of Thai-language tools.

Primary developer of these products:

- *Optical Character Recognition (OCR) software for Thai*
- *Thai word segmentation software*
- *Full-text indexing and searching software for Thai and English language, used widely in Thai governmental agencies, including the Parliament*
- *Document Capture & Indexing software for Krung Thai Bank (<http://www.ktb.co.th>) produced jointly with HP and FileNet*
- *Look Who's Clicking personalized Internet landing pages for direct mail, produced jointly with MindFire, Inc. of Irvine, CA, U.S.A., <http://www.lookwhosclicking.com>*
- *Statistical web page filtering service for parental control at <http://www.saijai.net>*

Director, Bangpo General Hospital (1995-)

Director, Modernform Integration Service (2005-)

Co-Founder, MFEC Public Company Limited (<http://www.mfec.co.th>)

Researcher, High-Energy Physics Group, California Institute of Technology/JPL-NASA (1989-1990)

Design and build scintillating high-energy particle detectors for the Superconducting Super Collider project.

Perform simulations of particle behaviors on several supercomputers.

Researcher, Applied Physics Group, California Institute of Technology/JPL-NASA (1990-1991)

Design and build secondary-electron-emission detector for time-of-flight heavy ion bombardment measurements.

Mathematical Consultant, First Quadrant Co., Pasadena, California, U.S.A. (1992)

Provide mathematical expertise for an exchange-rate prediction model for international currency trading program.

Researcher, Medical Imaging Group, UCLA, U.S.A. (1991-1995)

Research on radiological image compression and transmission on bandwidth-limited channels.

Lecturer in Physics, Mahidol University, Bangkok, Thailand. (1995-)

Lecture on introductory and fundamental physics for science students.

Lecture on mathematical methods of physics and numerical analysis.

Lecture on scientific research and presentation.

Publications and Presentations

Saipetch, P. (November, 1994). Lossless and high-fidelity archival storage of digital mammograms. Paper presented at the 1994 meeting of Radiological Society of North America, Chicago, USA

Saipetch, P. (November, 1994). Computerized contrast enhancement of size-specific features in mammograms. Paper presented at the 1994 meeting of Radiological Society of North America, Chicago, USA

Saipetch, P., Ho, B.K.T., Ma, M., Chuang, K.S., and Wei. Radiological image compression using wavelet transform with arithmetic coding. Proceedings of SPIE: Medical Imaging, Image Capture, Formatting, and Display. Vol. 2164, Newport Beach CA, Feb. 1994.

Saipetch, P., Ho, B.K.T., Panwar, R., Ma, M., and Wei, J. Application of wavelet transform with arithmetic coding in radiological image compression. IEEE Engineering in Medicine and Biology Sept. 1995.

Saipetch, P., Panwar, R., Ho, B.K.T., and Ma, M. Lossless image compression using wavelet-like filter banks and arithmetic coding. Proceedings of SPIE: Medical Imaging, Image Capture, Formatting, and Display. Vol. 2431, pp. 501-507, San Diego, CA, Feb. 1995.

Wei, J., Saipetch, P., Panwar, R., Chen, D. and Ho, B.K.T. Volumetric Image Compression by 3D Discrete Wavelet Transform (DWT) Proceedings of SPIE: Medical Imaging, Image Capture, Formatting, and Display. Vol. 2431, pp. 184-194, San Diego, CA, Feb. 1995.

Awards

1990 Lee Fisher Award in Physics, California Institute of Technology

1992 Chancellor Scholarship Award, University of California, Los Angeles

1994 Bailey Award, American Association of Physicists in Medicine

Lwin Moe

V.P. Tower
21/45 Soi Chawakun
Bangkok, Thailand 10400

1366 Hin Tha Myaing
Hyde Park (1) Road, Insein
Yangon, Myanmar

lwinmoe@gmail.com

Mobile Phone: (66)-05-848-6825

Objective

To help Myanmar improve its information technology infrastructure and help people to use this technology.

Education History

Indiana-Purdue University, Fort Wayne, IN, USA (graduated May, 2002)

- B.S., Information Systems.
- Minor in Mathematics.
- Graduated with highest distinction (Cumulative GPA: 3.85.)
- Dean's List.
- Who's who among students in American Universities 2002

Institute of Medicine (2), Yangon, Myanmar (universities closed 1996)

- Finished second year.

Employment History

June 2005-: IT Systems Manager, Center for Research in Computational Linguistics, Bangkok

- Purchase, install, and manage enterprise hardware and software systems
- Develop Burmese and Mon-language lexicographic and text corpus systems

Dec 2004-May 2005: Information Technology Consultant, Digital Divide Data Laos, Vientiane, Laos

- Provide in-class and on-job IT trainings to Digital Divide Data Laos (DDDL) staff (junior and senior management)
- Create a detailed IT curriculum for DDDL
- Provide a manual on the maintenance of DDDL's IT systems

Dec 2003-: Associate Professor, Myanmar Institute of Theology, Insein, Yangon, Myanmar

2003 Teaching part-time computer science and mathematics to Bachelor of Arts in Religious Studies (BARS) students.

2004 Helping the IT infrastructure at Myanmar Institute of Theology.

2005 Serving as a consultant in the areas of curriculum development and continuous development of the overall quality of BARS program.

2001-2002: Teaching and Research Assistant, Electrical Engineering Department, Indiana-Purdue University, Fort Wayne, Indiana, USA

1999 Assisted the head of the department of Electrical Engineering with preparing lecture notes, labs and projects for Web Engineering class.

2000 Researched relevant topics such as Web Engineering.

2001 Assisted students with their labs and projects.

1998-2001: Teaching and Research Assistant, Computer Science Department, Indiana-Purdue University, Fort Wayne, Indiana, USA

- Prepared lecture notes and labs for “Introduction to Computer Programming” and “Introduction to Computer Graphics Programming” classes.
- Answered questions from students and provided guidance with their lab and project work.
- Graded homework, labs and projects.
- Presented a research paper, “Visualization of Internet Search Results,” at the 14th annual undergraduate research conference at Butler University, Indianapolis, Indiana, USA.

1998-2002: Student Union Information Desk Student Worker, Indiana-Purdue University, Fort Wayne, Indiana, USA

Helped the setup and scheduling of the Student Union Office.
 Maintained IT equipment for student ID cards and student information system for the university.
 Helped new students to get familiar with the campus.

Extracurricular Activities

- Currently Involved in many activities of Friends of Burma (FOB), a committee to help the Myanmar Christians; Roles include advising, maintaining a clear communication flow with the FOB committee in the USA, coordinating tasks to produce best results, and producing reports.
- Helped people in the Myanmar community of Fort Wayne, Indiana with translation and interpretation needs.
- Went on a mission trip to Navajo reservation in New Mexico, USA to learn about their culture and help their families repairing roofs and volunteering at the local childcare center.
- Assumed the responsibility of public relations for Dr. Chit Maung Library, which is located at Gyogone, Insein.
- Co-founded Mon Bible School library in Mawlamyine, Mon State.
- Public Relations Coordinator for Karen Baptist Clinic at Seminary Hill in Insein.

Computer Experience

Operating System: Unix, Linux, Microsoft Windows.

Database: Oracle, Microsoft Access, MySQL.

Programming Languages: Java, C++, C, Perl, SQL, PHP, JavaScript, XML

Software: Microsoft Office, Maple, MiniTab, Adobe PhotoShop, PageMaker. GIMP

Website: <http://lwinmoe.friendsofburma.org>

APPENDIX 9. LICENSE TO USE THAI AUDIO LEXICON

Signed hard-copy originals of all letters are on file at CRCL Inc.

ThaiSoftware
Group

www.thaisoftware.co.th

วันที่ 4 พฤศจิกายน 2548

เรื่อง มอบสิทธิการใช้เสียงภาษาไทยของ ThaiSoftware
เรียน Mr. Doug Cooper
Center for Research in Computational Linguistics

บริษัท ไทยซอฟต์แวร์เอ็นเตอร์ไพรส์ จำกัด ขอขอบคุณที่ท่านให้ความสนใจในเสียงภาษาไทยของ ThaiSoftware ที่ใช้ในโปรแกรม ThaiSoftware Dictionary จึงมีความยินดีมอบสิทธิการใช้เสียงภาษาไทยของ ThaiSoftware เพื่อการศึกษาภายใน crcl.th.net สำหรับ SEAlang Lab Project ของหน่วยงาน Center for Research in Computational Linguistics เท่านั้น

ทั้งนี้ไม่มีขอบเขตเงื่อนไข ดังต่อไปนี้

1. อนุญาตให้ใช้สิทธิเสียงภาษาไทยของ ThaiSoftware เป็นเวลา 1 ปี ทั้งนี้หากท่านไม่ได้กระทำผิดเงื่อนไขใด ๆ ทางบริษัทฯ ก็ยินดีที่จะให้สิทธิการใช้เสียงภาษาไทยของ ThaiSoftware ในปีต่อไป อย่างต่อเนื่อง
2. ไม่สามารถโอน หรือ ขาย สิทธิการใช้เสียงภาษาไทยของ ThaiSoftware นี้ให้กับบุคคลหรือองค์กรอื่น ๆ ไม่ว่ากรณีใด ๆ
3. การนำเสียงภาษาไทยของ ThaiSoftware ไปใช้ จะต้องทำการป้องกันการ Copy หรือ Download เสียงภาษาไทยของ ThaiSoftware นี้อย่างเข้มงวด

ซึ่งหากท่านกระทำผิดจากเงื่อนไขดังกล่าวข้างต้น บริษัทฯ สงวนสิทธิ์ในการยกเลิกสิทธิการใช้เสียงภาษาไทยของ ThaiSoftware โดยไม่ต้องแจ้งให้ทราบล่วงหน้า

จึงเรียนมาเพื่อทราบ

ขอแสดงความนับถือ

มนัสนันท์ สุขุม

(นางนันทิสุข มณีรัตนะกุล)

ผู้อำนวยการบริหาร

Sales
FM-SAL-31-00

FM-จดหมายมอบสิทธิ
4/11/05

บริษัท ไทยซอฟต์แวร์ เอ็นเตอร์ไพรส์ จำกัด : 128/352 ชั้น 32 อาคารพญาไทพลาซ่า ถนนพญาไท แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพฯ 10400 โทร. 0-2612-0511 (อัตโนมัติ) โทรสาร. 0-2216-5153
บริษัท ไทยซอฟต์แวร์ (1999) จำกัด : 128/352, Phyathai Plaza Building, 32th Fl., Phyathai Road, Tung-Phyathai, Rajtavee, Bangkok 10400 Tel. 0-2612-0511(AUTO) Fax. 0-2216-5153

APPENDIX 10. LETTERS CONFIRMING COLLABORATION

DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER

FOREIGN SERVICE INSTITUTE

UNIVERSITY OF WISCONSIN-MADISON CENTER FOR SOUTHEAST ASIAN STUDIES / SEASSI

Signed hard-copy originals of all letters are on file at CRCL Inc.



REPLY TO
ATTENTION OF

DEPARTMENT OF THE ARMY

DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER
AND PRESIDIO OF MONTEREY
MONTEREY, CALIFORNIA 93944

Office of the Chancellor

November 4, 2005

Doug Cooper,
CRCL Inc.
820 Calle Pluma
San Clemente, CA 92673-2719

RE: 'SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition' project.

The Defense Language Institute Foreign Language Center confirms that it will collaborate in a pilot project to class-test the SEAlang Lab design and function.

- DLIFLC students and faculty will be directed to the SEAlang Lab facilities. These will be available on the Internet, and will not require any downloads or software installation by users.
- DLIFLC will periodically approve and distribute a survey form, supplied by CRCL, and intended to aid in ongoing project evaluation.
- DLIFLC and CRCL will hold a yearly workshop to *a)* evaluate student / instructor feedback; *b)* evaluate design of, and propose improvements to, the tools; *c)* document new uses or applications of the tools for the benefit of other SEAlang Lab collaborators.
- DLIFLC will be able to discuss technical or implementation issues with CRCL if the need arises.

This agreement is subject to the SEAlang Lab project receiving support under the U.S. Department of Education International Research and Studies Program (funding 10-2006 through 9-2009; approval to be announced by Q2 2006).

I look forward to utilizing SEAlang Lab in our Thai instructional program. It has significant potential for teaching the intricacies of this difficult language. Our program is of immense importance to U.S. defense policy, chiefly in the area of interoperability and other collaborative areas with the Royal Thai Army.

Sincerely,
/--signed--/
Neil Granoien, PhD
Vice Chancellor
Language Science & Technology



United States Department of State

Foreign Service Institute

*George P. Shultz National Foreign Affairs Training Center
Washington, D.C. 20522-4201*

November 10, 2005

Doug Cooper, Executive Director
CRCL Inc.
820 Calle Pluma
San Clemente, CA 92673-2719

Dear Mr. Cooper:

I am writing on behalf of the Foreign Service Institute (FSI) to confirm plans to collaborate with CRCL in beta-testing and evaluating the tools proposed under the 'SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition' project. It is understood that this project is being submitted to the U.S. Department of Education International Research and Studies program, and if approved will begin in October, 2006 and last through September 2007. This pilot will be at no cost to FSI.

FSI's role/responsibilities will consist of:

- the tools, which are Web-accessible, and require no software installation by users, will be made available at no cost and on a volunteer basis to FSI students and faculty;
- the FSI will distribute a project evaluation form (provided by CRCL) that invites anonymous feedback;
- a yearly workshop will be held at FSI, at a mutually convenient time, at CRCL's expense, to:
 - o discuss how to use the tools most effectively,
 - o discuss design and possible improvement of the tools,
 - o evaluate student and teacher feedback, and
 - o document innovative tool applications for sharing with other collaborators;
- interested FSI faculty and staff will be able to discuss technical issues related to tool implementation with CRCL staff at any time;

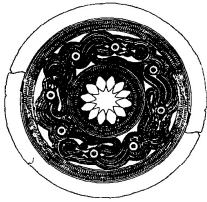
We understand that the SEAlang Lab tools should be considered to be experimental and development-in-progress until the project is completed. We stipulate that, in the unlikely event that no one at FSI expresses interest in taking part in this project, or that the volunteers withdraw from it, the project would be terminated at the request of FSI, at no cost to FSI.

If the project is successful in demonstrating that use of the tools you have developed contributes to efficient and effective learning of the languages and cultures of Southeast Asia, it will be a significant achievement.

Thank you for making the tools available to FSI for testing.

Sincerely yours,

Colleen E. Lewis
FSI/GSO/Acquisitions Officer

**UNIVERSITY OF WISCONSIN—MADISON****Center for Southeast Asian Studies**

207 Ingraham Hall
1155 Observatory Drive
Madison, WI 53706-1397
Telephone: 608/263-1755
Fax: 608/263-3735
E-mail: seasia@intl-institute.wisc.edu

November 14, 2005

Doug Cooper, Executive Director
CRCL Inc.
820 Calle Pluma
San Clemente, CA 92673-2719

Dear Mr. Cooper:

I am writing on behalf of the University of Wisconsin-Madison Center for Southeast Asian Studies to confirm that we will collaborate with CRCL Inc. on this project:

SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition in Complex-script Languages

As the host (2000-2009) of the Southeast Asian Studies Summer Institute (SEASSI, sponsored by a consortium of 14 North American universities), CSEAS will invite faculty to integrate the SEAlang Lab tools into the Institute's teaching and teacher-training programs, as well as into its academic-year language programs.

As with the SEAlang Library, all tools will be made available to CSEAS and SEASSI students and faculty, who will be invited to meet with CRCL periodically to discuss how to use the tools most effectively, discuss design and improvement, evaluate student and teacher feedback, and document applications to share with other collaborators.

CSEAS looks forward to continuing to work together with CRCL Inc. on these useful and important projects.

Sincerely,

A handwritten signature in black ink, appearing to read "R. Anderson Sutton". The signature is fluid and cursive.

R. Anderson Sutton, Director

APPENDIX 11. LETTERS OF SUPPORT

COORDINATOR, INTERAGENCY LANGUAGE ROUNDTABLE

PRESIDENT AND VICE-PRESIDENT, COUNCIL OF TEACHERS OF SOUTHEAST ASIAN LANGUAGES

EXECUTIVE DIRECTOR, NATIONAL COUNCIL OF LESS COMMONLY TAUGHT LANGUAGES

EXECUTIVE DIRECTOR, LINGUISTIC DATA CONSORTIUM

DIRECTOR, SOUTHEAST ASIAN STUDIES SUMMER INSTITUTE

Signed hard-copy originals of all letters are on file at CRCL Inc.



The Interagency Language Roundtable

2005: The Year of Languages in the Federal Government

November 12, 2005

ILR Steering Committee

Frederick H. Jackson, *Foreign Service Institute*
ILR Coordinator

Beth Mackey, *Department of Defense*
Co-Chair of Testing Committee

Christina N. Hoffman, *Foreign Service Institute*
Co-Chair of Testing Committee

Lea Christiansen, *Intelligence Language Institute*
Co-Chair of Training Committee

Douglas F. Gilzow, *Foreign Service Institute*
Co-Chair of Training Committee

Maria Brau, *Federal Bureau of Investigation*
Co-Chair of Translation & Interpretation Committee

Paul Hopper, *Department of State*
Co-Chair of Translation & Interpretation Committee

Scott McGinnis, *Defense Language Institute, Co-Chair of Special Interest Group for CASL*

Cynthia Riquelme, *Defense Intelligence Agency, Co-Chair of Special Interest Group for CASL*

Gale Holdren, *Fulbright-Hays Programs, Department of Education*

Edward McDermott, *International Education, Department of Education*

Laura Murray, *Center for Advanced Study of Language, U. of Maryland*

Doug Cooper, Executive Director
CRCL Inc.
820 Calle Pluma
San Clemente, CA 92673-2719

Dear Mr. Cooper:

As Coordinator of the federal Interagency Language Roundtable, I am writing to express the ILR's appreciation for your excellent presentation on September 30 to the first ILR Plenary of this academic year.

The level of interest in the kind of work you are doing was reflected in the large and attentive audience for your presentation, by the questions that you were asked and by the number of people who came up and spoke to you after the presentation.

As an applied linguist and language educator with considerable experience in Southeast Asia and with Southeast Asian languages, I am personally and professionally tremendously excited about several of the features of the tools that you have developed and that I had the opportunity to work with during your visit.

First and foremost is what you are doing to make languages with extremely complex non-roman orthographies truly accessible to the learner. By providing orthographic word-breaks when needed, and glosses of words or phrases, you provide windows into the languages. The fact that any online text in such languages as Thai, Khmer, Lao, Burmese or Shan—including today's newspaper texts—can be manipulated by your tools is especially exciting.

Also of considerable interest are your tools to increase reading fluency and text processing by, e.g., increasing or decreasing the size of the font with a simple toggle, changing the font itself on the fly, and displaying running text at speeds chosen by the learner to develop word and phrase recognition.

Perhaps most exciting, however, as I think of my own professional work with these languages (and others like them) is what you have accomplished to permit the understanding and analysis of text in context. Because your work for each language is based upon extremely large corpora of texts from many different genre, and because you have solved the problem of providing online searchable access to an extraordinarily broad range of monolingual and bilingual (and trilingual!) dictionaries and other reference materials, including

<http://www.govtilr.org>

some that were published more than 100 years ago, your tools are, if anything, even more useful to the professional language analyst, descriptive linguist or language course developer than they are to the beginning learner. Thanks to your work, it will soon be possible for an analyst of any of these languages to pull up instantly all examples in the corpora (or a defined set of the corpora) of words or strings of words—even when they are discontinuous—and then analyze in detail even the finest nuances of their use and usage. It is also possible to compare the meanings and use of cognates and borrowings across a range of languages. These kinds of capabilities have been available for English and most Western European languages for some time, and of them have been available for character-based languages like Japanese, Chinese and Korean, but those of us working on mainland Southeast Asian languages have not been so fortunate. With the tools you are developing, that promises to change.

It is not only Southeast Asian languages, however, that lack the kinds of capabilities that you have developed for Thai, Khmer and Lao. To the best of my knowledge (and the knowledge of several experts with whom I have spoken) there exist few if any comparable tools for the languages of South Asia that are written in Sanskrit-derived devanagari orthographies (e.g., Hindi, Nepali, Bengali, Malayalam, Tamil, Sinhala) or for the many Middle Eastern and Central Asian languages that are written in orthographies based on Arabic (e.g., Arabic, Persian, Urdu, Pashto, Dari.) Nowhere are your kinds of tools more needed than in the study and analysis of texts in these languages.

Let me once again express my appreciation, on behalf of the ILR, for coming such a great distance to talk with us about your work, and also my admiration for the quality of that work and for the commitment you have demonstrated to thinking “outside the box” in seeking solutions to problems that were previously considered intractable.

Sincerely,

Frederick H. Jackson, Ph.D.
ILR Coordinator

<http://www.govtilr.org>



COTSEAL

Council of Teachers of Southeast Asian Languages

November 12th, 2005

To Whom It May Concern:

As the President and Vice President of the Council of Teachers of Southeast Asian Languages (COTSEAL), and on behalf of COTSEAL members, we firmly support the Sealang Lab project.

We are writing this letter to underscore the project's significance to the Southeast Asian Language teaching profession. We teach quite difficult languages, but the revolution in computer-assisted language learning stops in its tracks when it encounters the Thai, Khmer, Burmese and Lao scripts.

That is why the Sealang Lab software is so exciting. COTSEAL members are extremely committed to making instruction more relevant to our students' diverse needs. We can already use the Internet to find authentic materials, but the vernacular press, especially in Thailand and Cambodia, is very difficult for learners to understand. It is intensely frustrating that we have no reading or writing tools that can help a student read a daily local newspaper, or write a letter to the editor. The Sealang Lab software will address these basic, and very important, problems.

Moreover, unlike most language software projects, the Sealang Lab takes the approach of helping the student without telling him or her what to study, and of helping the professor without telling him or her how to teach. This is very welcome because it gives COTSEAL members the tools we need to investigate new methodologies for teaching Southeast Asian languages. Such research is crucial for our continuing professional development within the field, and it is critical to our ability to meet national needs in foreign language education in the long term.

If COTSEAL as an organization or we as Southeast Asian language teachers can assist you in any way, please do not hesitate to let us know.

Sincerely yours,

Bac Hoai Tran
COTSEAL President
E-mail: bact@calmail.berkeley.edu

Leo Paz
COTSEAL Vice President
E-mail: lpaz@ccsf.edu



November 10, 2005

To Whom it May Concern:

I am writing to support the CRCL Inc. proposal: **SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition in Complex-Script Languages.**

I have read the proposal, seen a live demonstration presented by Project Director Doug Cooper, and discussed the project with Mr. Cooper.

The issues raised by the SEAlang Lab proposal regarding technology in teaching and learning less commonly taught languages are extremely important for American students and faculty.

Students who must learn the difficult scripts addressed by this proposal would certainly benefit enormously from supporting software. But there has been little development of such tools – taken singly, these least-commonly taught languages have the fewest students, and naturally get the least support. Yet, as the proposal points out, many languages have shared or similar scripts, even when the languages themselves are not related. When taken as a whole, the Arabic, Devanagiri, and Southeast Asian scripts this proposal will help address impact a very substantial number of students and teaching programs.

But it is of even greater concern that our teaching faculty have so few opportunities to pursue research in these areas. It is a simple fact that much modern research in second language acquisition relies heavily on the availability of software tools; and it is equally apparent that most less common language teaching faculty have neither the time nor the opportunity to create them. It is not necessary for language teachers to become programmers – it is, however, necessary to recognize that funding for technical support has become a critical element of research.

The SEAlang Lab project provides an unusual and effective example of the way in which we can solve practical student problems, while at the same time enabling serious faculty research into the pedagogy of second language acquisition. As Executive Director of the National Council of Less Commonly Taught Languages (NCOLCTL), I urge you to give this proposal the strongest possible consideration for support.

Sincerely,

Antonia Folarin Schleicher
Executive Director, NCOLCTL



LINGUISTIC DATA CONSORTIUM

3600 Market St., Suite 810, University of Pennsylvania, Philadelphia, PA 19104-2653 USA • Tel: (215) 898-0464 • Fax: (215) 573-2175 • email: ldc@ldc.upenn.edu

Tuesday, November 15, 2005

Doug Cooper
Center for Research in Computational Linguistics, Bangkok

Dear Doug,

I am writing to express my support for your project: *The SEAlang Lab: Assistive Technology for Reading, Writing, and Vocabulary Acquisition in Complex-Script*. I am familiar with your work from several encounters. In preparation for my duties as discussant of your presentation to the Interagency Language Roundtable (ILR), I reviewed the SEAlang site and your papers. I also had the opportunity to attend your ILR presentation and demonstration and to meet subsequently to discuss possible joint work. Finally, during a recent harvest of resources to support human language technology development in Thai, I was reminded of just how many have your name on them.

As I commented at the ILR meeting your innovative research addresses important gaps in current knowledge. There are certainly numerous raw resources, lexicons and technology components such as word segmenters in Thai. Malay is less well served but not bereft. Other groups, including LDC, are working on resource collection and sharing. However, there remains a significant gap in the application of language resources and human language technologies to the problems of language learning and teaching. This problem is only exacerbated by the increasing emphasis on authentic language in teaching and the rapidly growing need for analysts, commentators and scholars who understand foreign cultures and their media. At a time when external needs and current teaching practice force language learners to grapple with authentic language earlier in their programs and government linguists to wrestle with growing heaps of information, we find too little technological support. The workbenches you describe, and for which you have already demonstrated working components, will support the future learner as s/he comes to terms with the challenges of the information age.

As you know your project is well aligned with work LDC is currently doing for the Department of Education. Both projects focus on authentic language and the developments of specific skills such as reading. They differ appropriately in the linguistic issues they address: complex morphology of Arabic, complex orthography and compounding in Thai. In both cases the graphical user interface is simply a shell for the more interesting and complex linguistic technologies embedded in it. In particular, allowing users to input text and queries using loose phonetic transcription coupled with fuzzy matching will prove extremely helpful not only to beginning learners who have not yet mastered the orthography but also to advanced learners working with rapid conversational speech that has proven a challenge for humans and systems alike in the case of the CallHome project.

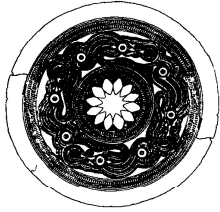
I would like to underscore the opportunities for collaboration that we discussed and that would be facilitated if your proposal were funded. LDC would be willing to distribute resources you have created or would create, giving them a broader audience among the 1820 organizations worldwide that use our data. In our work for the REFLEX: Less Commonly Taught Languages project, we have already acquired several million words of news text and several hundred thousand words of Thai-English parallel text with distribution rights. These could become building blocks of your prepared learning resources. Finally, I believe we could exchange tools and ideas both virtually and through a regular series of meetings, perhaps semi-annual and coordinated with major conferences that we both attend.

In summary I support your research, look forward to seeing your latest innovations take root and hope for an opportunity to further advance our collaboration.

Sincerely,

A handwritten signature in black ink, appearing to read 'C. Cieri'.

Christopher Cieri
Executive Director, Linguistic Data Consortium



UNIVERSITY OF WISCONSIN—MADISON

Center for Southeast Asian Studies

207 Ingraham Hall
 1155 Observatory Drive
 Madison, WI 53706-1397
 Telephone: 608/263-1755
 Fax: 608/263-3735
 E-mail: seasia@intl-institute.wisc.edu

November 10, 2005

Doug Cooper, Executive Director
 CRCL Inc.
 820 Calle Pluma
 San Clemente, CA 92673-2719

To Whom it May Concern:

I am writing as Director of the Southeast Asian Summer Studies Institute (SEASSI) to support the "SEALang Lab: Assistive Technology" project, which was discussed at the SEASSI Executive Committee meeting in Chicago this spring, and demonstrated to a group that included myself, the SEASSI Language Director, and interested faculty and staff at our intensive session in Madison this summer.

SEASSI is a cooperative program of fourteen Southeast Asian Studies Centers, including eight National Resource Centers. We offer the only intensive summer Southeast Asian language program in the U.S. We serve a broad range of students who are preparing for overseas study, work, or field research, including undergraduates, graduate students, and professionals. Also, many SEASSI participants each summer are Southeast Asian Americans who already have relatively high speaking and listening skills in the language of their heritage. Therefore, SEASSI has many students who, already having intermediate level proficiency, are poised to gain advanced levels of proficiency. The tools proposed in SEALang Lab will greatly help them make the leap to advanced proficiency.

The software tools demonstrated as part of the SEALang Lab will provide an immense service to SEASSI, and to our member Centers. Although the project is initially directed at Thai, it is clear from the demonstration I saw that the "Workbench" software and methodology it provides will be readily adaptable to other languages. There is no question that the American Southeast Asian language programs that sponsor SEASSI will use these tools, because they address real, practical needs at all stages of instruction.

This is exactly the kind of innovative instructional material that the International Research and Studies program should sponsor, and I do not hesitate in giving it the strongest possible recommendation.

Yours sincerely,

Dr. Michael Cullinane,
 Director, Southeast Asian Studies Summer Institute

Budget Narrative

Budget Narrative

Attachment 1:

Title: Pages: Uploaded File: **9048-Mandatory_irs05-budget.pdf**

Budget Narrative, SEAlang Lab

U.S. Department of Education, International Research and Studies SEAlang Lab Budget, Years 1- 3, October 1 2006 - September 30, 2009				
		Federal		
		Year 1	Year 2	Year 3
1. Personnel				
	Project Director, 40%	\$55,848	\$58,082	\$60,405
	Lexical resources manager, 50% of 9-month salary	\$21,750	\$22,620	\$23,525
	<i>Personnel Subtotal</i>	<i>\$77,598</i>	<i>\$80,702</i>	<i>\$83,930</i>
2. Fringe				
	Personnel only, 34%	\$26,383	\$27,439	\$28,536
3. Travel				
	Meetings with FSI, DLI, SEASSI, CSEAS	\$5,000	\$5,000	\$5,000
4. Equipment		\$0	\$0	\$0
5. Supplies				
	Desktop PCs for project staff, \$1,200 X 3	\$3,600	\$0	\$0
6. Contractural		\$0	\$0	\$0
7. Construction		\$0	\$0	\$0
8. Other				
	Senior curriculum consultant, \$60 X 160 hours	\$9,600	\$9,600	\$9,600
	Linguist consultants, \$40 X 320 hours	\$12,800	\$12,800	\$12,800
	Software consultant, \$500 X 15 days	\$7,500	\$7,500	\$7,500
	Native speaker data developer, \$1,000 X 12 months	\$12,000	\$12,000	\$12,000
	IT manager, 40% X \$2,000 X 12 months	\$9,600	\$9,600	\$9,600
	Office manager, 40% X \$1,000 X 12 months	\$4,800	\$4,800	\$4,800
	Outside evaluator, \$3,000 (including expenses)		\$3,000	
	Dedicated web server, fees \$200 X 12 months	\$2,400	\$2,400	\$2,400
	Office rent & expenses, 40% X \$1,500 X 12 months	\$7,200	\$7,200	\$7,200
	<i>Other Subtotal</i>	<i>\$65,900</i>	<i>\$68,900</i>	<i>\$65,900</i>
9. Total Direct Costs		\$178,481	\$182,041	\$183,366
			<i>Total direct costs</i>	<i>\$543,888</i>
10. Indirect Costs, 10%		\$0	\$0	\$0
11. Training Stipends		\$0	\$0	\$0
TOTAL		\$178,481	\$182,041	\$183,366
			<i>Total Federal costs</i>	<i>\$543,888</i>

Notes: The budget is nearly identical from year to year, except for \$3,600 in first-year equipment costs, a \$3,000 contract for outside evaluation in year 2, and a 4% yearly increase in base salary rates.

1. Personnel

We are requesting a 40% appointment for the Project Director. This position requires skill in the analysis, design, and management of technical projects, a high degree of competence in computational linguistics and software development, and extensive experience in working with Thai. Compensation follows published NSF guidelines for similar work (USED does not publish guidelines). Note that the Project Director is only compensated for contracted project work, and is not paid for serving as Executive Director of CRCL Inc.

We request a 50% appointment for the Lexical Resources Manager, whose base rate matches his academic-year salary at the University of Wisconsin-Madison.

2. Fringe Benefits

CRCL Inc. works closely with faculty and staff of the University of Wisconsin-Madison, both as an employer and as a subcontractor. The CRCL Inc. fringe rate of 34% equals the University of Wisconsin fringe rate for equivalent positions.

3. Travel

We are requesting \$5,000 per year in travel costs. This will cover domestic travel, lodging, and per diem for yearly visits to our collaborators at the Foreign Service Institute (Arlington, Virginia), Defense Language Institute (Monterey, California), and Center for Southeast Asian Studies / Southeast Asian Studies Summer Institute (Madison, Wisconsin). Any remaining funds will support conference attendance.

4. Equipment

No equipment expenses are included.

5. Supplies

We request support for three mid-range office desktop PCs, equivalent to Dell OptiPlex Desktop (\$1,200 X 3, total \$3,600), including monitors, software, and network support.

6. Contractual

Not applicable.

7. Construction

Not applicable.

8. Other

Consultants and non-US citizen employees in our Bangkok office are paid the following negotiated fees for the lifetime of the contract, and do not receive fringe benefits.

Consulting rates are typical for highly specialized work of this type. Fees paid to the native speaker data developer, IT systems manager, and office manager reflect local compensation in Bangkok and are considerably lower than equivalent U.S. costs.

- 160 hours / year at the rate of \$60 / hour for the Senior Curriculum Consultant.
- 320 hours / year at the rate of \$40 / hour for Linguist consultant(s).
- \$3,000 (including expenses) in year two for the outside project evaluator.
- 15 days / year at the rate of \$500 / day for the Software Consultant.
- \$12,000 / year for the native speaker data developer.
- \$9,600 (40% of \$24,000 / year) for the Bangkok IT systems manager.
- \$4,800 (40% of \$12,000 / year) for the Bangkok office manager.

The project also requests support for dedicated Web server services (equivalent to one Dell PowerEdge 1850, which is a standard mid-range rack-mount server), including Internet access fees. A dedicated server with adequate speed and memory (2 gigabytes), and very low-latency network access will make a substantial difference in the quality of SEALang Lab services, providing essentially instantaneous feedback on predictive completion features of the Writer's Workbench, and minimal delays on the audio features of the Reader's and Vocabulary Workbenches. Note that server software support will be provided by the Bangkok IT systems manager. The rate of \$200 per month is typical for the configuration and performance we require.

The project requests a 40% share of our office expenses (including utilities and telephones, which total approximately \$1,500 per month). Remaining costs are paid by CRCL Inc. as part of its cost-share obligation to the SEALang Library project.

9. Total Direct costs

\$543,888 provided by the Federal government.

10. Indirect costs

None requested.

11. Training Stipends

Not applicable.

12. Total Costs

\$543,888 provided by the Federal government.