INTERNET RESOURCES FOR SOUTHEAST ASIAN LINGUISTICS

Gwyn Williams
Yuphaphann Hoonchamlong
Thammasat University

Introduction

We are living in an era of internationalization in which Information Technology (IT) [1] is playing an increasingly important role in people's lives, especially those who are involved in the acquisition, exchange or dissemination of information. Academics make up a large proportion of such people. With the rapid growth of the Internet, the *inter*connected *net*work of computer networks, more and more people can communicate and retrieve information from any corner of the world with increasing ease and speed. Apart from greater accessibility to information, contacts with other experts in the same and related fields and interested individuals are facilitated.

There are three commonly used means of getting information from the Internet that are relevant to our discussion, namely, e-mail, mailing lists, and World Wide Web.

1) **E-Mail (Electronic Mail)** is the electronic equivalent of the post office. Instead of sending a letter, a message is sent in electronic form between computers and reaches its destination in seconds.

2) A **Mailing List** is the e-mail equivalent of a newsletter and discussion forum. When subscribed to a mailing list one is able to receive up-to-date information and participate in ongoing discussion on specific areas of interest. LINGUIST List, with over 6,000 subscribers, is one of the largest electronic mailing lists in the world.

3) **World Wide Web (WWW)** is a means of accessing electronic information of all kinds in the form of text, pictures, sound, and movies from many interlinked databases. One can instantly obtain further details on a topic although the information may be at different geographical locations simply by selecting a key word on the computer screen.

The academic Internet connection in Thailand is provided by ThaiSarn (*Thai Social/Scientific, Academic and Research Network*), operated as part of NECTEC (*National Electronic and Computer Technology Center*), Ministry of Science, Technology and Environment.

---

[1] Defined as "The acquisition, processing, storage and dissemination of vocal, pictorial, textual and numerical information by microelectronics-based combination of computing, telecommunications and video" (Longley & Shain 1989).

We would like to introduce you to some of the resources relating to Southeast Asian linguistic studies that are available via the Internet.

Southeast Asian Resources at Australian National University
WWW Address: http://coombs.anu.edu.au/WWWVL-AsianStudies.html

ANU maintains an extensive and up-to-date database for Social Sciences & Pacific and Asian Studies. Information relating to Southeast Asia is organized by region and country. There are also links to Asian and Southeast Asian resources at other universities, such as the University of Wisconsin-Madison, USA, Asian Institute of Technology (AIT), Thailand, and NECTEC. The table of contents is shown in part in Figure 1. One can access data screens for individual Southeast Asian countries either through the main Asian menu or through the Southeast Asian menu. We have included Laos and Thailand as examples of the latter.

Figure 1. Asian Studies at Australian National University

Asian Studies
* What's New in WWW Asian Studies (ANU,Australia)
* Asia Internet Report (silkroute.com,Singapore)
* Clearinghouse for Social Sciences Subject-Oriented Bibliographies (ANU,Australia)

Asia-Pacific Studies Internet Resources - Central Database
* ANU-Asia-WWW-Gopher-News-L dbase
* Search the InfoSeek Net Search (www2.infoseek.com,USA)

Asian Studies - Index
Asia-Pacific region as a whole * Middle East & Caucasus * Central Asia * South Asia * SouthEast Asia * Australia & New Zealand * Pacific Ocean *
East Asia
Afghanistan * Armenia * Assyria * Australia * Azerbaijan * Bahrain * Bangladesh * Bhutan * Brunei * Burma * Cambodia * Chechnya * China * East Timor * Formosa * Georgia * Hong Kong * India * Indonesia * Iran * Iraq * Israel * Japan * Jordan * Kazakhstan * Korea (North) * Korea (South) * Kurdistan * Kuwait * Kyrgyzstan * Laos * Lebanon * Macau * Malaysia * Maldives * Mongolia * Myanmar * Nepal * New Zealand * Oman * Pakistan * Palestine * Papua New Guinea * Philippines * Qatar * Russian Far East * Saudi Arabia * Siberia * Singapore * Sri Lanka * Syria * Taiwan * Tajikistan * Thailand * Tibet * Turkey * Turkmenistan * United Arab Emirates * Uzbekistan * Vietnam * West Bank * Yemen

ASIAN CONTINENT AS A WHOLE
[...]
    SOUTHEAST ASIA
        o South East Asia Archives (U.Calif,Irvine,USA)
        o South East Asia Newsletter (U.Calif,Irvine,USA)
        o Southeast Asian Resources (U Wisconsin-Madison,USA)
        o South East Asian Ministers of Education Organization - SEAMEO (AIT,Thailand)
        o Major Libraries with Southeast Asian resources (U Wisconsin-Madison,USA)
        o Thai Yunnan Project Newsletter (ANU & NECTEC,Thailand)
        o Review of Bitnet/Internet Lists for Southeast Asia (ANU,Australia)
        o The Commonwealth of Learning - S.E.Asia Region Page (www.col.org)

o Asie du Sud-Est (tertius.ina.fr,France)
o Southeast Asia History Archives (neal.ctstateu.edu,USA)
o SEASCALINFO! - Southeast Asian Studies at Cal (U.Calif.,Berkeley,USA)
o Institute of Southeast Asian Studies (ISEAS, Singapore)
[many listed Asian countries deleted]
Laos
Status: Republic
Capital: Vientiane (Viangchan)
o Link to Asian Studies - Laos [8 links] (ANU,Australia)
[...]
Thailand
Status: Kingdom
Capital: Bangkok (Krung Thep)
Networking Resources
o Asian Institute of Technology Server (AIT,Thailand) Information Resources
o Thailand - WWW Virtual Library (NECTEC,Thailand)
o ATHENS - Australia Thailand Higher Education Network Service (ANU,Australia)
o ThaiOnline - WWW resource of the National Thai Studies Centre (ANU,Australia)

Southeast Asian Resources at the University of Wisconsin-Madison
WWW Address: http://www.library.wisc.edu/resources/SEAsia/htmls/menu04.htm

ANU provides a direct link to Southeast Asian resources at the University of Wisconsin-Madison, USA. The colorful opening menu, as viewed with a graphical browser program, is shown in Figure 2. Information on each country can be accessed by clicking on the country flag. Since the preparation of this paper, we have found that East Timor has recently been added.

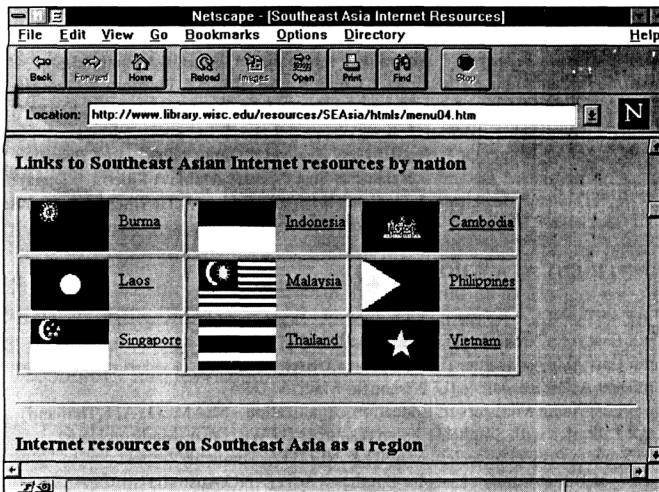Figure 2. Wisconsin - Gateway to Southeast Asia: Opening Menu

Figure 3 shows what is available on Southeast Asia at Wisconsin University, besides country information. In addition to access to library resources and records at Wisconsin itself, one can obtain information about the University's Center for Southeast Asian Studies, as well as follow links to libraries and resources at other locations.

Figure 3. Wisconsin Southeast Asian Resources Page: Table of Contents

UW Campus Libraries:
Southeast Asia in campus libraries
Bibliographies, Indexes, and Abstracts:
Current Periodical Subscriptions at U.W.
Microfilm Collections at U.W.
New in the U.W. General Library System
U.W. South & Southeast Asia Video Archive
Social Weather Bulletin Index
Rand research abstracts
U.W. Center for Southeast Asian Studies:
Information about the Center and its faculty
Center Publications
Internet Resources:
Descriptions and links to Internet resources on Southeast Asia
Descriptions and links to other major libraries with Southeast Asian resources
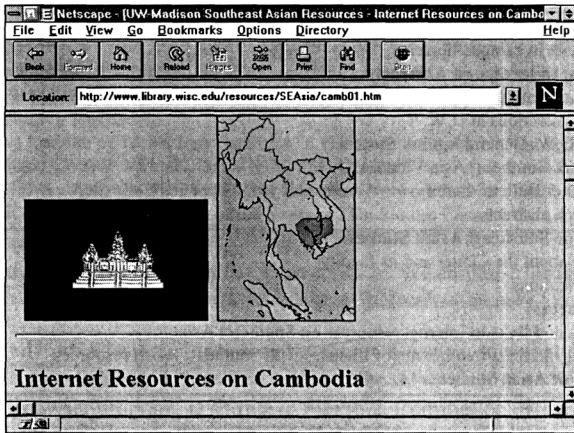Association of Asian Studies

Figure 4 shows more about the microfilm collections at the University of Wisconsin. Microfilms include newspapers from the Philippines, Thailand and *Indonesia, old Malay publications, Indonesian texts from World War II and later,* and political documents relating to the Philippines and Thailand.

Figure 4. Wisconsin Microfilm Resources On Southeast Asia

Microfilm Collections on Southeast Asia
Newspapers/Serials: Memorial Library owns a large collection of microfilmed newspapers. We currently subscribe to Philippine Daily Inquirer (Philippines), Prachachat Thurakit (Thailand), and Kompas (Indonesia). In addition to these subscriptions, the Memorial Library regularly acquires newspaper titles on an as needed basis. Newspapers are cataloged and can be found using MadCat.

Colloquial Malay: The Colloquial Malay collection presently contains over twenty microfilm reels of monographs and serials in Malay from late 19th and early 20th century Indonesia. *Sometimes referred to as "low Malay," or "Peranakan Malay," the collection represent an effort to* develop a comprehensive resource that details the rise of the vernacular press in the Dutch East Indies. In addition to a rich source of data for research on a variety of social, political, literary and linguistic topics such as national language development, the rise of nationalism, the rise of the vernacular press, studies of ethnic relations, popular culture studies, and the transition from classical Malay syair (poetic) and hikayat (prose) forms to western literary genres of the novel and short story. Individual items are accessible through the General Library's catalog, MadCat. A bibliographical databases is under development. Collection located in Microforms and Media Collection, 4th Floor Memorial Library.

For each country in Southeast Asia, Wisconsin provides unusually good graphics, including the flag and a map of each country. One of the more complete and interesting Pages is that for Cambodia, as shown in Figure 5. As though to demonstrate just how quickly World Wide Web sites are updated, this Page was revised in the week this paper was being prepared.

Figure 5. Wisconsin Internet Resources on Cambodia: Opening Menu

Most of this Page deals with the recent history and politics of Cambodia, but some newspapers are available. Of more interest to linguists, under Language and Literature, there is a link to the Ethnologue Database.

Ethnologue Database
WWW Address: http:// www.sil.org/ethnologue

The Ethnologue is an electronic version of *Ethnologue: Languages of the World* (Grimes 1992). Statistical data and other information on 6,500 languages are provided. Languages of Southeast Asia are very well covered. Information is organized by country, language name, and language family. The Ethnologue is totally unique: nowhere else is it possible to obtain such extensive information so easily and so quickly.

Figure 6 gives the country profile for Cambodia. We find that, in addition to general facts, 18 languages are spoken in Cambodia. The file also provides recent population and language figures, when available.

Figure 6. Ethnologue Database: Cambodia

```
Cambodia
Continent       Asia
Population       6,592,000 (1990 WA)
Remarks          State of Cambodia. Formerly Kampuchea, Khmer Republic. Literacy rate 50%
                 (1990). Information mainly from Thomas and Headley 1970; Voegelin and
                 Voegelin 1977; Wurm and Hattori 1981. Data accuracy estimate: B, C
Blind population  40,000 (1982 WCE)
Religion         Theravada Buddhist 90%, secular 4%, traditional religion 3%, Muslim 3%
Languages
BRAO             5,000 in Cambodia (1988 govt. figure); 18,000 total in Laos and Cambodia
                 (1984) over 90 in USA (1990); 5 in France; 30,000 together with Krung 2 and
                 Kravet (1981 Wurm and Hattori)
CHAM, WESTERN    195,000 in Cambodia (1988 govt. figure); 10,000 in Malaysia; 4,000 in
                 Thailand; 3,000 or more in USA; 1,000 in France; 100 in Saudi Arabia; 213,000
                 total or more
CHINESE, MANDARIN 300,000 in Cambodia (1985 Gunnemark and Kenrick); 885,000,000 total
                 Mandarin (1991 WA)
CHONG            5,000 possibly in Cambodia; 500 possibly in Thailand (1981 SIL); 8,000 total
                 (1981 Wurm and Hattori)
KHMER, CENTRAL   5,932,800 in Cambodia, about 90% of the population (1990); 320,000 in
                 Thailand; 700,000 in Viet Nam (1985); 50,000 or more in France; 50,000 in
                 USA; 10,400 in Laos; 7,063,200 total
KRAVET           3,000 (1988 govt. figure); 30,000 together with Brao, Kru'ng (1981Wurm and
                 Hattori)
KRUNG 2          (30,000 together with Brao, Kravet; 1981 Wurm and Hattori)
KUY              15,000 in Cambodia (1988 govt. figure); 200,000 in Thailand (1991UBS);
                 650,000 total (1981 Wurm and Hattori)
LAMAM            1,000 (1981 Wurm and Hattori)
MNONG, CENTRAL 19,000 in Cambodia (1988 govt. figure); 90,000 total (1981)
PEAR             1,300 (1988 govt. figure)
SAMRE            200 (1981 Wurm and Hattori)
SAOCH            500 (1981 Wurm and Hattori)
SOMRAY           2,000 (1981 Wurm and Hattori)
STIENG           70,000 total (1981 Wurm and Hattori); 48,000 in Viet Nam (1973 SIL)
SUOY             200 (1981 Wurm and Hattori)
TAMPUAN          13,500 (1988 govt. figure)
VIETNAMESE       300,000 in Cambodia (1985 Gunnemark and Kenrick); 59,000,000 total (1991
                 WA)

Part of the Ethnologue Database
```

The Ethnologue is also organized by individual language. Basic information about Central Khmer, spoken in Cambodia, is shown in Figure 7. As the Ethnologue points out, the varieties of Khmer spoken in Cambodia and Thailand are quite distinct.

Figure 7. Ethnologue Database: Central Khmer

```
KHMER, CENTRAL (CAMBODIA)
Country            Cambodia
Language code   KMR
Continent          Asia
Alternate language names   CAMBODIAN
Genetic affiliation       Austro-Asiatic, Mon-Khmer, Eastern Mon-Khmer, Khmer
Geographical region       In USA: Long Beach and Orange County, California; Seattle,
                   Washington; Oregon; North Dakota; New York; Rhode Island;Wisconsin; Texas
Population         5,932,800 in Cambodia, about 90% of the population (1990); 320,000 in
                   Thailand; 700,000 in Viet Nam (1985); 50,000 or more in France; 50,000 in
                   USA; 10,400 in Laos; 7,063,200 total
Bible printings   1954
New Testament printings   1929-1984
Printings of whole books of Bible   1899-1933
National language   Yes
Remarks            Distinct from Northern Khmer of Thailand
Grammar typology        SVO
Total speakers   7,063,200
Also in            Viet Nam, Thailand, France, USA, Laos
```

Khmer is a member of the Mon-Khmer language family. The Ethnologue gives a listing of the many branches and languages in this family (Figure 8). Khmer is listed as part of Eastern Mon-Khmer, along with Bahnaric, Katuic, and Pearic.

Figure 8. Ethnologue Database: Mon-Khmer Language Family

```
AUSTRO-ASIATIC, MON-KHMER
  * Aslian
    + Jah Hut
      o JAH HUT (Malaysia, Peninsular)
    + North Aslian
      o Chewong (1 language)
      o Eastern (4 languages)
      o Tonga (1 language)
      o Western (3 languages)
    + Senoic
      o LANOH (Malaysia, Peninsular)
      o SABUM (Malaysia, Peninsular)
      o SEMAI (Malaysia, Peninsular)
      o SEMNAM (Malaysia,
        Peninsular)
      o TEMIAR (Malaysia, Peninsular)
    + South Aslian
      o BESISI (Malaysia, Peninsular)
      o SEMAQ BERI (Malaysia,
        Peninsular)
      o SEMELAI (Malaysia,
        Peninsular)
      o TEMOQ (Malaysia, Peninsular)
```

```
  * Eastern Mon-Khmer
    + Bahnaric
      o Central Bahnaric (5 languages)
      o North Bahnaric (14 languages)
      o South Bahnaric (7 languages)
      o West Bahnaric (11 languages)
    + Katuic
      o East Katuic (14 languages)
      o West Katuic (10 languages)
    + Khmer
      o KHMER, CENTRAL
        (Cambodia)
      o KHMER, NORTHERN
        (Thailand)
    + Pearic
      o Eastern (1 language)
      o Western (5 languages)
  * Monic
    + MON (Myanmar)
    + NYAHKUR (Thailand)
  * Nicobar
    + Car
      o NICOBARESE, CAR (India)
```
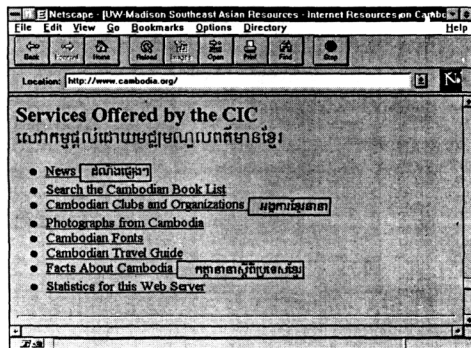
+ Chowra-Teressa
  o CHAURA (India)
  o TERESSA (India)
+ Great Nicobar
  o NICOBARESE, SOUTHERN (India)
+ Nancowry
  o NICOBARESE, CENTRAL (India)
+ Shom Peng
  o SHOM PENG (India)
* **Northern Mon-Khmer**
+ Khasian
  o AMWI (India)
  o KHASI (India)
  o PNAR (India)
+ Khmuic
  o Khao (2 languages)
  o Mal-Khmu' (7 languages)
  o Mlabri (1 language)
  o Xinh Mul (3 languages)
+ Mang
  o MANG (Viet Nam)
+ Palaungic
  o Eastern Palaungic (7 languages)
  o Western Palaungic (15 languages)

* **Palyu**
  + PALYU (China)
* **Viet-Muong**
  + Chut
    o AREM (Viet Nam)
    o MAY (Viet Nam)
    o RUC (Laos)
    o SACH (Viet Nam)
  + Cuoi
    o HUNG (Viet Nam)
    o PONG 2 (Laos)
    o TUM (Laos)
  + Muong
    o BO (Laos)
    o KHA TONG LUANG (Laos)
    o MUONG (Viet Nam)
    o NGUON (Viet Nam)
    o PONG 1 (Laos)
  + Unclassified
    o PAKATAN (Laos)
    o PHON SUNG (Laos)
    o THAVUNG (Laos)
  + Vietnamese
    o VIETNAMESE (Viet Nam)

Another link takes us to the Austro-Asiatic language family, of which Mon-Khmer is one of two branches, the other being Munda.

From the Wisconsin Cambodian Page (Figure 5) there are links to other informative sites, such as the Cambodian Information Center Home Page. The opening screen features an image of Angkor Temple. Figure 9 shows one of the following screens. This Page is notable for its use of the Khmer script. We also learned that we were the 17,130th visitor to this site.

Figure 9. Cambodian Information Center

From these pages one can access maps, data screens, and recent issues of Cambodian English language newspapers, such as *The Cambodia Times*.

A pleasant surprise for anyone surfing through the Wisco.isin Gateway database (Figure 2) is to be found in the Philippines Page (Figure 10). There is a great deal of interesting information on Philippine languages and scripts with many accompanying images.

Figure 10. A Philippine Leaf: Table of Contents

[...] * **Laguna Copperplate Inscription**. A document from 900 A.D. that threatens to upset our long-held theories of Philippine history.
* **Literacy in Pre-Hispanic Philippines**. When Miguel de Legazpi came to Manila, he found that almost everybody could read and write. This prompted the friars to publish a book in the native script in 1593, forty-seven years before the first book was published in the United States.
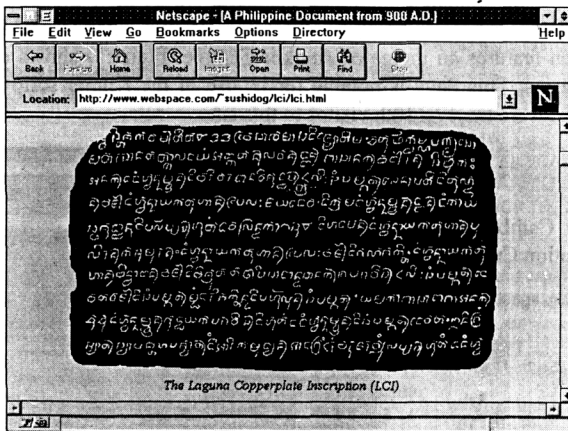* **Extinction of a Philippine Script**. Why did the Tagalog script become extinct just a little over a century after the Spaniards came?
* **The Tagalog Script**. The writing system that was in place when the Spaniards arrived.
* **Our Living Scripts**. Not until almost the end of the Spanish era did it become known that remote mountain groups had maintained their literacy in scripts similar to the Tagalog script. They are still in use today.
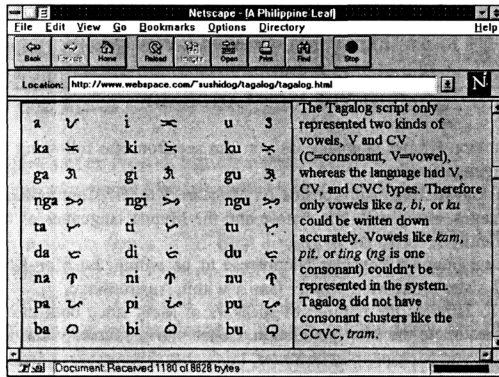
A discussion by Hector Santos of the recently discovered Laguna copperplate inscription is to be found here (Figure 11). The inscription, found in 1989 and containing a date 900 AD, is in Kavi script and appears to contain a mix of Sanskrit, Old Tagalog, Old Javanese, and Old Malay, which could make this text contemporary with Angkor and Champa civilizations.

Figure 11. The Laguna Copperplate Inscription

The Laguna Copperplate Inscription (LCI)

A portion of the Tagalog script listed in the menu above (Figure 10) can be seen in Figure 12 below.
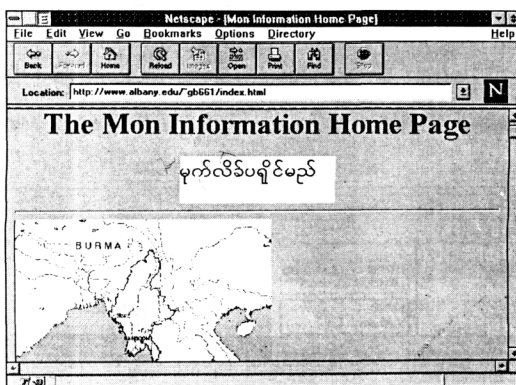
Figure 12. Philippine Leaf: Talalog Script

The Philippine site is certainly worth a visit and will provide hours of interesting reading, as will the entire University of Wisconsin-Madison site, which is unrivaled.

While we are on the theme of Southeast Asian writing systems and scripts, there are several other useful sites which provide samples of and information on Southeast Asian scripts and, most useful of all, computer fonts. One very good site for Mon language is to be found at *http://www.albany.edu/~gb661/index.html* (Figure 13).

Figure 13. The Mon Information Home Page

Quite sophisticated linguistic information on Mon is available at this site (Figure 14). For example, the Mon-Khmer language family, of which Mon is a member, is available, this time in tree form (Figure 15). The tree is very large, so we can shown it only in part.

Figure 14. The Mon Language

THE MON LANGUAGE

Mon is a Mon-Khmer language, as you can see from the following tree, a chart of the Austro-Asiatic language family.

The best known linguistic cousins of Mon are Khmer (Cambodian) and Vietnamese. Mon-Khmer languages, along with Nicobarese and the Munda languages of India together form the Austroasiatic family.
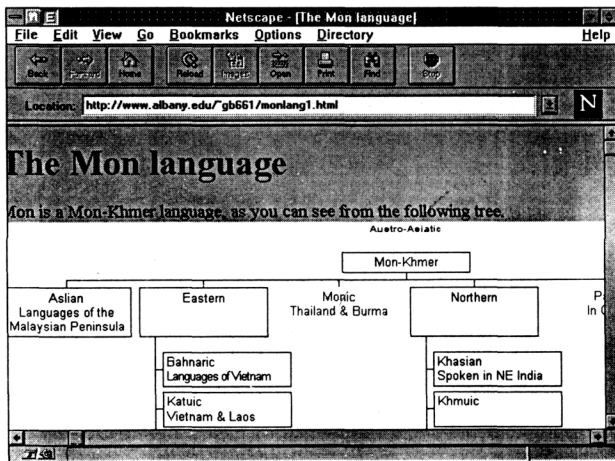
Note: Some linguists consider Nicobarese to be within the Mon-Khmer family, while others consider it a sister to the Mon-Khmer family within Austroasiatic.

It's hard to estimate the number of speakers of Mon, since both the Thai and Burmese governments underestimate the Mon population. Conservative sources claim that there are about 800,000 speakers. Some Mon people, however, think that the number is likely to be five or six times as great.

Mon is written in an Indic-based alphabet which is derived from Pali. The Mon writing system was adapted for use in the Burmese language. You can see a sample of this alphabet at the Burmese Font page. For more information on Mon grammar, follow one of these links:

    * Mon phonetics and phonology
    * Mon morphology
    * Mon syntax
    * A bibliography of Mon linguistics

Figure 15. The Mon Language: Family Tree

If we follow the link to Mon phonetics and phonology, we will find a brief description of Mon register, including a review of the debate about the nature of register in Mon (Lee 1983, Diffloth 1985, Thongkum 1987) and the vowels and consonants (also available in IPA characters).
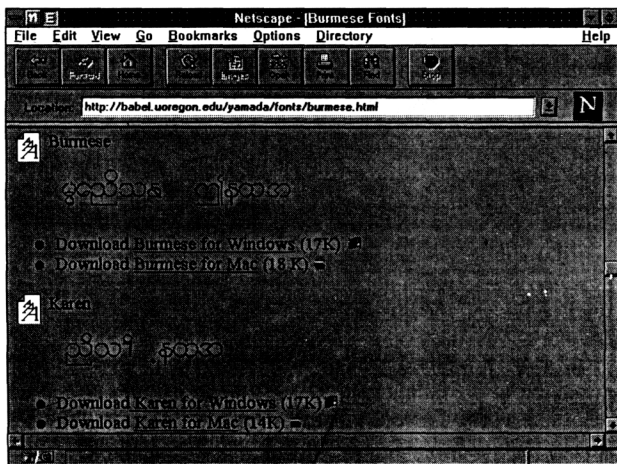
This Mon page, already of exceptional quality and one of the most detailed for any Southeast Asian language, bears a notice that it is still under construction. It is most certainly one site to keep an eye on and one worthy of emulation.

Yamada Language Center Non-English Font Archive
WWW Address: http://babel.uoregon.edu/yamada/team.html

Let us turn now to fonts. The best location to find fonts is the Yamada Language Center, University of Oregon, which maintains an archive of Macintosh and Windows compatible fonts for languages with non-English (non-Roman) scripts. All these fonts are public-domain, which means they are freely available to be downloaded to one's own computer. Southeast Asian fonts to be found here are Burmese, Karen, Lao, Thai, and Vietnamese. Samples of each font may be viewed on-screen before they are down-loaded. Figure 16 shows a sample of the Burmese fonts that are available. Figure 17 shows some fonts for Vietnamese.
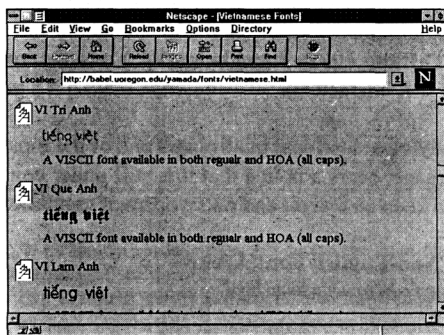
Figure 16. Burmese fonts

Figure 17. Vietnamese fonts

ThaiOnline
WWW Address: http://online.anu.edu.au/thaionline/

Let us return now to our point of departure: Australian National University. ANU electronic resources for Thailand include ThaiOnline, the last entry in Figure 1, provided by the National Thai Studies Centre of ANU. Figure 18 shows the opening screen.
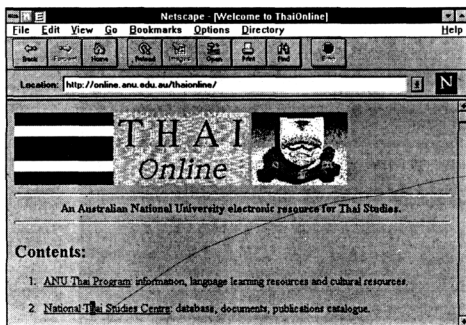
Figure 18. ThaiOnline

In addition to information on Thai language study and culture, images, and the Thai Studies Centre at ANU, ThaiOnline provides links to the Ethnologue Database and LaoOnline, an ANU National University electronic resource for Lao language and culture studies.

Besides accessing databases via the World Wide Web, the Internet provides another important means to obtain information in the form of special interest groups or mailing lists. Mailing lists are a forum for discussion, exchange of ideas, and fielding questions. There are three such lists that are directly relevant to Southeast Asian linguistics. All three are operated from NECTEC, Bangkok.

SEALANG-L
SEALANG-L is an international scholarly forum and central archive for the discussion, documentation, and dissemination of information on the languages spoken in Southeast Asia. The scope of discussion includes formal features of these languages, as well as issues related to theory, research, history, relationship,

sociolinguistic situation, language planning, and teaching. There are over 200 subscribers, many of whom are well-known experts in languages of the region. This year-old list has already proven its value with lively discussions on the nature of stop consonants in Thai, register, and the origin of the word "Siam". [2]

THAITEACH

THAITEACH was initially proposed at the July 1995 meeting of the Council for the Advancement of the Study of Thai (CAST) to provide a channel of communication and discussion for those interested in Thai language teaching for non-native speakers. Its purpose and scope is to provide an international forum for the discussion and dissemination of information concerning the teaching of Thai to non-native speakers, for example, discussion and assessment of audio-visual and written materials for teaching Thai, Thai language programs, etc. [3]

NLPASIA-L

NLPASIA-L provides an international scholarly forum and central archive to facilitate and co-ordinate the documentation, research, discussion, and dissemination of information on the computational processing of languages in Asia. Main topics include: natural language processing; computational linguistics; large scale corpora; speech processing; and related fields. There are about 200 subscribers. [4]

Conclusion

Information Technology is playing an increasingly important and unique role in the modern world of scholarly research and collaboration. We encourage scholars of Southeast Asian languages to make full use of available information technology to explore the many resources which have been provided by academic instituitions and individual researchers. Modern technology provides unique and powerful tools to make and to maintain avenues in the exchange of ideas and observations and to bridge gaps in direct communication and contact.

References

Longley, Dennis and Michael Shain. 1989. Macmillan Dictionary of Information Technology. London: The Macmillan Press.

Grimes, Barbara F. (ed.). 1992. Ethnologue: Languages of the World. Twelfth edition. SIL.

---

[2] For further information, contact the list owner, Mr. Gwyn Williams, at <gwyn@ipied.tu.ac.th>.

[3] For further information, contact the list owner, Dr. Yuphaphann Hoonchamlong, at <yui@ipied.tu.ac.th>.

[4] For further information, contact the list owner, Mr. Gwyn Williams, at <gwyn@ipied.tu.ac.th>.