Catch and release:
caring feeding sharing
of comparative data

Doug Cooper Center for Research in Computational Linguistics
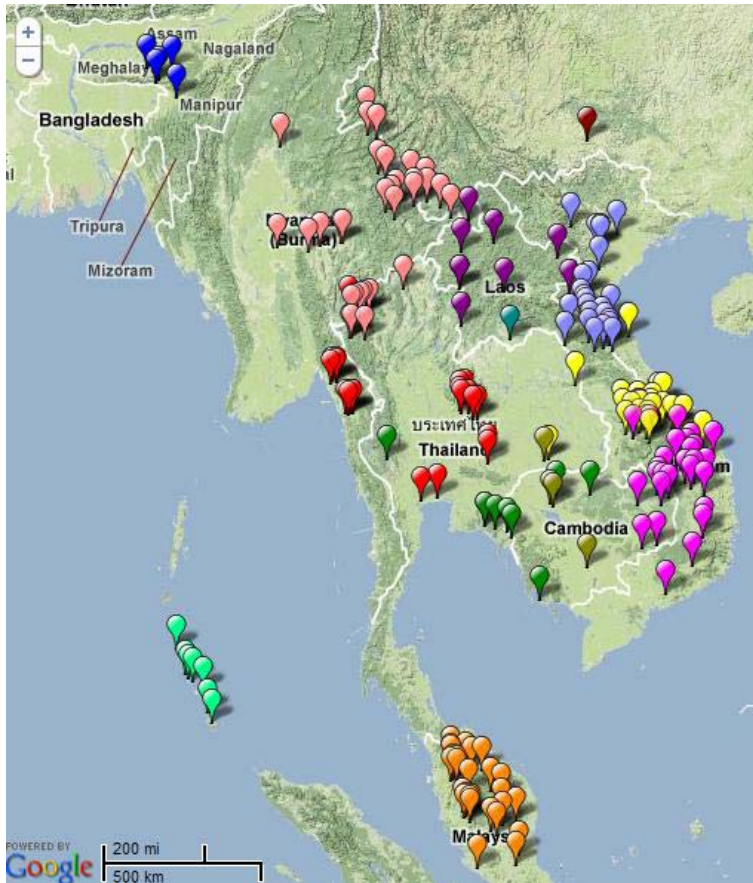
today

one	problems
two	solutions
three	lessons
four	questions

Part 1: Problems

**Who says
what why
when
where
& how ?**

**Mapping
and managing
comparative
data in
the MKLP**

It's not just about language.

it's about data…

we're great at writing it (field notes, unpublished work).
we're great at publishing it (books, articles, theses).
we're great at warehousing it (libraries, archives).
we're great at describing it (metadata, OLAC).

but not so good at

access
navigation
illumination

and websites…

very good at presenting results

not so good for testing ideas

we need to turn websites into worksites

and tools…

many systems to pour data into:

Elan, WordCorr, WordSurv…

few tools to get to know data with

and human-computer interaction…

specifying searches

organizing answers

we're not building ordinary dictionaries

we need extraordinary controls

and collaboration…

many conferences
many publications

little
anonymous
unilateral
cooperation

enabled by open access to data
controlled access to analysis

wilensky

and interoperability…

Internet

as

sprachbund

using standardized markup (TEI, GOLD, ...)

and/or Web APIs

a.k.a. "linguistic area, convergence area, diffusion area or
language crossroads"

and accessibility.

lexicography

epigraphy

endangered languages

helping ask
expert questions
without demanding
expertise in each area

data

tools

websites

human-computer interaction

accessibility

collaboration

interoperability

collaboration care

data feeding

websites sharing

human/computer interaction of

tools comparative

accessibility language

interoperability data

Part 2: Solutions

basic dictionary layout

basic dictionary controls

IPA query specification

clear [pbɓpʰ] [eɛæɾiəɐʌɤ] [ɲnŋ] | sav

clear entry	Bilabial	Lab'dent	Apical	Retro
Nasal V U	m m̥	ŋ ŋ̊	n n̥	ɳ ɳ̊
Plosive V U I	b p ɓ		d t ɗ	ɖ t
(Aspirate	Select all voiced in this row		tʰ	
Fricative S+	ɸ	f	θ s ʃ	ʂ
Appxt V U	w w̥	v(ʋ) ʋ̥	l l̥	r r̥

Choose languages | **Choose IPA characters** | **Choose sources** | Help | History | Issues | Map results

Build a query *or* one bracketed allophone set   [abc] = a *or* b *or* c, *but* (ab|cd) = ab *or* cd. See the **MKED Tutorial and Cookbook** for more deta

clear [pbɓpʰ] [eɛæɾiəɐʌɤ] [ɲnŋ] | save   save as [saved]   **V** = any vowel, **X** = any consonant

clear entry	Bilabial	Lab'dent	Apical	Retro	Palatal	Velar	Uvular	Phary'l	Glottal
Nasal V U	m m̥	ŋ ŋ̊	n n̥	ɳ ɳ̊	ɲ ɲ̊	ŋ ŋ̊	N N̥		
Plosive V U I	b p ɓ		d t ɗ	ɖ t	ɟ c ʄ	g k ɠ	ɢ q		ʔ
(Aspirated) A	pʰ		tʰ		cʰ	kʰ			
Fricative S+	ɸ	f	θ s ʃ	ʂ	ç	x	χ	ʕ ħ	h
Appxt V U	w w̥ v(ʋ) ʋ̥		l l̥	r r̥	j(ɥ) ɥ̊	ɰ ɰ̊	ʀ ʀ̥		

	Front		Center Back		
High	i	y	ɨ ʉ	ɯ	u
U? C LB	ɪ	ʏ		ᵿ	ʊ
	e	ø	(ɜ) ə		o
	ɛ æ	œ	ɐ ʌ		ɔ
Low			a	ɑ	ɒ

Tables are derived from IPA, but reflect typical Southeast Asian practice.   Vowel layout is similar to Pullum & Ladusaw (1996), pp. 298-299.

Search these works ...

Wat2009 Pacoh (Watson, Richard)
Khasic
Sin1906 Khasi (Singh, U. Nissor)
Wei1975 War (Weidert, Alfons K.)
Khmeric
Dha1978 Surin Khmer (Chantrupanth
Hea1997 Khmer (Headley, Robert K.
Khmuic
Fil2009 Mal (Filbeck, David)
Lin1974 Khmu (Lindell, K.)
Ris1995 Mlabri (Rischel, Jorgen)
Suw2002 Khmu (Suwilai Premsrirat)
Mangic
Edm1995 Bolyu (Edmondson, Jerold
Monic
Dif1984 Mon (Diffloth, Gérard)

include or exclude authors and works

| Choose languages | Choose IPA characters | **Choose sources** | Help | History | Issues | Map results |

Search these works ...

Include all data
General
Shorto MKCD (Sho2006)
Huffman 20 MK (Huf1971)
Aslian
Bis1994 Kensiu (Bishop, Nancy M. a
Kru2004 Semelai (Kruspe, Nicole D.)
Mea1998 Temiar (Means, Natalie)
Bahnaric
Ban1979 Bahnar (Banker, John; Ban
Cro2004 Tampuan (Crowley, James
Smi2000 Sedang (Smith, Kenneth D.)
Katuic
Chi1978 Ir (Chinda Kommala)
Cos1971 Katu (Costello, Nancy A.)

Exclude these works ...

Don't exclude anybody
All but ancestors (unless 'full')
General
Shorto MKCD (Sho2006)
Huffman 20 MK (Huf1971)
Aslian
Bis1994 Kensiu (Bishop, Nancy M. a
Kru2004 Semelai (Kruspe, Nicole D.)
Mea1998 Temiar (Means, Natalie)
Bahnaric
Ban1979 Bahnar (Banker, John; Ban
Cro2004 Tampuan (Crowley, James
Smi2000 Sedang (Smith, Kenneth D.)
Katuic
Chi1978 Ir (Chinda Kommala)

Search these languages ...

No Language Restriction
RECONSTRUCTIONS
proto Katuic (1395 items)
proto Mon (620 items)
proto Mon-Khmer (8160 items)
proto Monic (685 items)
proto Nyah Kur (703 items)
proto Pearic (149 items)
proto Pong (760 items)
proto Semai (49 items)
proto Vietic (1238 items)
proto Wa-Lawa (544 items)
proto Waic (544 items)
LANGUAGES
Acehnese (58 items)

Reanalysis ...

Allow non-conflicting additions
Allow all additions
Allow original authors only
Allow author-approved
Allow additions by...
Cooper (Coo2007)
Exclude additions by...
Cooper (Coo2007)

Original search target(s):
phonetic transcription: "[pbɓpʰ] V[nn̥tddʼtʰθsʃll̥]" (*as*
(?:p|b|ɓ|p(?:h|ʰ))
(?::|ʔ|a|e|i|o|u|ɔ|ə|ɯ|ɛ|ʊ|ɑ|ɨ|ʉ|ɪ|ʏ|ø|ɘ|ɵ|ɤ|œ|ɜ|ɞ|ʌ|æ|ɐ|ɑ|ɒ)+
(?:n|(?:hn)|(?:n(?:l̥)))|t|d|dʼ|t(?:h|ʰ)|θ|s|ʃ|l|(?:hl)|(?:l(?:l̥))))
gloss: "fly" (*as* **fly, flies, flew, flied, flown, flying**)

expanded queries

SEAlang Mon-Khmer
Languages Project
Home / Dictionary / Database

SEARCH Normal Exact
Reset

[pbɓpʰ]V[nn̥tddʼtʰθsʃll̥]

IPA

_____ **Orth**

fly **Text** (☑
expand)

Original search target(s):
phonetic transcription: "[pbɓpʰ] V[nn̥tddʼtʰθsʃll̥]" (*as*
(?:p|b|ɓ|p(?:h|ʰ))
(?::|ʔ|a|e|i|o|u|ɔ|ə|ɯ|ɛ|ʊ|ɑ|ɨ|ʉ|ɪ|ʏ|ø|ɘ|ɵ|ɤ|œ|ɜ|ɞ|ʌ|æ|ɐ|ɑ|ɒ)+
(?:n|(?:hn)|(?:n(?:l̥)))|t|d|dʼ|t(?:h|ʰ)|θ|s|ʃ|l|(?:hl)|(?:l(?:l̥))))
gloss: "fly" (*as* **fly, flies, flew, flied, flown, flying**)
language(s): "**212 languages.**"
"**V**" matches any vowel.
Normal search - ignoring **B C D** modifiers, and allowing standard
vowel / consonant approximations. Including both reconstructions
and citations in search.
25 items found

single items

Original search target(s):
   phonetic transcription: "k.*m" (*as* k.*m)
   language(s): "212 languages."
Normal search - ignoring **B C D** modifiers, and allowing standard vowel / consonant approximations. Inc
both reconstructions and citations in search.
1490 items found

MON-KHMER RECONSTRUCTIONS (27 ITEMS)

ASLIAN BRANCH (26 ITEMS)

ˣ ˢ **kam** *kam* *rat trap (noun)* (Temiar) MEA1998:C:1255

ˣ ˢ **kantam pǝnagŋ** *field crab (compound)* (Kensiu) BIS1994:C:482

ˣ ˢ **kantam** *crab* (Kensiu) BIS1994:C:481

ˣ ˢ **karom** *area underneath a house; underneath, below (noun)* (Semelai) KRU2004:C:49

ˣ ˢ **kem** *military camp (noun)* (Semelai) KRU2004:C:741

ˣ ˢ **kem** *kem* *a squirrel trap (noun)* (Temiar) MEA1998:C:1379

ˣ ˢ **kennem** *kennem* *mint (noun)* (Temiar) MEA1998:C:1427

ˣ ˢ **khlǝm** *odor (noun)* (Semelai) KRU2004:C:791

ˣ ˢ **khoːm** *to sit, be quiet* (Mah Meri) SKEAT & BLAGDEN 1906 S 223 SHO2006:C:1427-5 via variant?

ˣ ˢ **khoːm** *to sit, be quiet* (Mah Meri) SKEAT & BLAGDEN 1906 S 223 SHO2006:C:1427-6 via variant?

ˣ ˢ **kim** *kim* *we will, indicates the future tense when used as a marker with the exclusive pronouns kanek / kandek (we) (future pronoun marker)* (Temiar) MEA1998:C:1495

ˣ ˢ **klam** *to carry on shoulder* (Semelai) SKEAT & BLAGDEN 1906 C 36 SHO2006:C:1406-2

ˣ ˢ **kobm** *to be half-open (eyes)* (Kensiu) BIS1994:C:580

| X | S | *ɲaʔ | *house* | (proto Waic) | DIF1980:R:ʔ23-1 |

dummy root (proto Samtau) DIF1980:R:ʔ23-2

nyaʔ_ *house* (P'uman) FER1971:C:DIF1980~2

nha *house* (Kem Dègne) LEF1892:C:DIF1980~3

ɲàʔ *house* (Samtau) HAR1976:C:DIF1980~25

nya *house* (Tai Loi) SCO1900:C:DIF1980~31

dummy root (proto Wa-Lawa-La) DIF1980:R:ʔ23-3

*ɲaʔ *house* (proto Wa-Lawa) DIF1980:R:ʔ23-4

ɲɨaʔ *house* (Lawa [North]) DIF0000:C:DIF1980~2

nī-ö *house* (Wa) MIL1931:C:DIF1980~3

nyeh *house* (Wa) ANT1911:C:DIF1980~5

ɲɨaʔ *house* (Lawa [Ban Pa Phae]) SUR1976:C:DIF1980~6

ñïea *house* (Lawa [L'up]) SCH1976:C:DIF1980~8

ñeʔ³ *house* (Wa [Thung Va]) LUC1965:C:DIF1980~9

nyīiex *house* (Wa [Kawa]) KAW1958:C:DIF1980~10

nyeh *house* (Wa [Bible]) YOU1934:C:DIF1980~11

nyen *house* (Wa) DAV1909:C:DIF1980~11

niüa, nia⁵ *house* (Wa) DRA1907:C:DIF1980~12

ɲeʔ *house* (Wa [South]) HAR1976:C:DIF1980~26

ɲɨaʔ *house* (Lawa [Bo Luang]) MIT1972:C:DIF1980~37

ɲɨaʔ *house* (Lawa [Ban Phae]) MIT1972:C:DIF1980~38

ɲɨaʔ *house* (Lawa [Umphai]) MIT1972:C:DIF1980~39

cha(t) *house* (La) DAV1909:C:DIF1980~12

nya *house* (En) SCO1900:C:DIF1980~29

nya *house* (Son) SCO1900:C:DIF1980~30

nha *house* (Kentung-Wa) SCO1900:C:DIF1980~32

nya *house* (K'ala) HAR1927:C:DIF1980~5

nya *house* (Khalo) FLA1970:C:DIF1980~5

family trees

rat, mouse (proto Mon-Khmer) SHO2006:R:93

  *kn₁[i]ʔ rat, mouse (proto Mon-Khmer [A]) SHO2006:R:93.A
    kniʔ kni(') rat, mouse (Old Mon) SHO2006:C:93-1
    nɔeʔ rat, mouse (Mon) SHO2006:C:93-2
    (k)nai rat, mouse (Kui) SHO2006:C:93-3
    kəneːi rat (Stieng) SHO2006:C:93-4
    ɗe rat (Sre) SHO2006:C:93-5
    neː rat (Stieng [Biat]) SHO2006:C:93-6 and ɗɛː < Sre ?
    kənɛː rat (Chrau) SHO2006:C:93-7
    kənɛː rat (Bahnar) SHO2006:C:93-8
    kaneːy rat (Jeh) SHO2006:C:93-9
    kəneː rat (Halang) SHO2006:C:93-10
    kənéʔ rat, mouse (Khmu [Yuan]) LIN1974:C:SHO2006~93-11
    khnai mouse (Khasi) SHO2006:C:93-12
    kĕni mouse (Sakai [Central]) SHO2006:C:93-13
  *knɛɛ rat (proto Katuic) SID2005:R:572
    knaj~naj rat, mouse (Kui) PRA1978:C:SID2005~572-1
    knaj rat (Souei) FER1974A:C:SID2005~572-2
    kanaj rat, mouse (Bru) THE1980:C:SID2005~572-3
    kanʌj rat, mouse (Bru) MIL1976:C:SID2005~572-4
    kanaj rat (Sô) GAI1985:C:SID2005~572-5
    kunəj rat (Katang [T]) SID2005:C:572-6
    kunɛɛ wild game (pig) (Pacoh) WAT1979:C:SID2005~572-7

and more family trees

basic database layout

⊙ Mon... ❌ | ⊙ SEAlang... | 🐫 Lingua::... | ⊙ Mahidol ... | SKE Sketch E... | ⊙ Franklin ... | ⊙ huffman... | 📄 Free de... | 📊 Spreads...

SEAlang Mon-Khmer Languages Project
Home / Etymological Dictionary / Database

(choose)	(sort)	(display)	
books by author	◉ IPA	◉	**Custom dictionary**
books by branch	○	HTML	**Save file**
all languages	orthography	○ text	**Reset**
branch+language	○ gloss	○ XML	
	○ ID	○ JSON	

Antisdel, C. B.
1911 **Elementary studies in Lahoo, Akha (Kaw), and Wa languages**. Journal of the Burma Research Society 1.1:41-64.
show all 92 entries :: build custom dictionary

Banker, John; Banker, Elizabeth; Siu Mo
1979 **Bahnar Dictionary: Plei Bong - Mang Yang Dialect**. Summer Institute of Linguistics, California. xvi, 202 p.
show all 6283 entries / view source :: build custom dictionary

Bishop, Nancy M. and Peterson, Mary M.
1994 **Kensiw Rhyming List**, TUSIL=LRDP Thammasat University, Bangkok, Thailand.
show all 1635 entries / view source :: build custom dictionary

Chantrupanth, Dhanan
1978 **Khmer (Surin) - Thai - English Dictionary**, Bangkok: Chulalongkorn University.
show all 3859 entries :: build custom dictionary

Chinda Kommala
1978 **Ir Wordlist**, unpublished.
show all 180 entries :: build custom dictionary

Costello, Nancy A.
1971 **Ngữ-vựng Katu: Katu Vocabulary**. Vietnam Montagnard Language Series 5, Saigon. Department

About the Mon-Khmer Languages Database

The Mon-Khmer Languages Project's primary resources are this **Languages Database**, devoted to preservation and sharing of language and lexical resources, and a companion **Etymological Dictionary** built to support work in comparative and historical linguistics.

New capabilities are being added in the week of October 12. Please excuse any temporary disruptions in service, and join us at **ICAAL 4** in Bangkok, October 29-30, 2009 to discuss new and proposed features.

Sources of Language Data

The Mon-Khmer Languages project is collecting data on some 150 languages. Our primary sources are:

published dictionaries As a rule, smaller dictionaries designed for linguistic reference and research are digitized word-for-word, while larger language reference works may be condensed.

unpublished notes and lexicons Much valuable work on languages that are highly endangered, or are spoken in areas that are now relatively inaccessible for political reasons, can only be found in archival storage.

contributed data Many researchers have data that was insufficiently analyzed for their original research purposes, but is nevertheless of value for comparative linguistics.

Attribution / naming / glossing

Every data item is uniquely labeled with its author (three-letter abbreviation), year of publication, an original identifying label or number if provided, and possibly a SEAlang "C" (for "citation") or "R" (for "reconstruction") number.

Some etymological and comparative dictionaries do not necessarily gloss all of the citations that follow a reconstruction or vice versa. In these cases an interpolated gloss will be completely parenthesized. "Dummy" entries (marked as **clusters**) are sometimes added to help group etymologically related terms; e.g. when the underlying dictionary has been developed along semantic rather than comparative lines. Typical IPA and gloss values are interpolated from the cluster members.

Some of the more obscure works cited on the left have been extracted from larger comparative dictionaries. These items have compound identifiers that show both the original work, and the borrower's usage, e.g.
The1980:C:Sid2005~25-3 indicates that Sidwell 2005's entry 25-3 quotes or paraphrases Theraphan 1980.

database controls

SEAlang Mon-Khmer Languages Project
Home / Etymological Dictionary / Database

(choose)	(sort)	(display)	
books by author	⦿ IPA	⦿ HTML	**Custom dictionary**
books by branch	○ orthography	○ text	**Save file**
all languages	○ gloss	○ XML	**Reset**
branch+language	○ ID	○ JSON	

Banker, John; Banker, Elizabeth; Siu Mo
1979 **Bahnar Dictionary: Plei Bong - Mang Yang Dialect**. Summer Institute of Linguistics, California. xvi, 202 p.
show all 6283 entries / **view source** :: **build custom dictionary**

☐ Bahnar, all 6283 items
- ☑ Bahnar (Golar), 1899 items
- ☐ Bahnar (Kontum), 647 items
- ☑ Bahnar (Pleiku), 3737 items

Bishop, Nancy M. and Peterson, Mary M.
1994 **Kensiw Rhyming List**, TUSIL=LRDP
Thammasat University, Bangkok, Thailand

extracted dictionaries

Custom Dictionary
Language:Wa, Source:Ant1911   92 entries

GLOSS	IPA	LANGUAGE (DIALECT)	ID
a La, a Lawa, a Wa	hsiveh	Wa	Ant1911:C:Dif1980−19
all, out of, exhausted	oet	Wa	Ant1911:C:Dif1980−49
arrive	hoet	Wa	Ant1911:C:Dif1980−51
bamboo	o:	Wa	Ant1911:C:Dif1980−0
big	ding	Wa	Ant1911:C:Dif1980−36
bird	hsim	Wa	Ant1911:C:Dif1980−67

combined dictionaries

Custom Dictionary
Language:Chong, Source:Huf1971
Language:Chong, Source:Huf1985
Language:Chong, Source:Isa2007a
Language:Chong, Source:Isa2007b
Language:Chong, Source:Nop2003
Language:Chong, Source:Por2001
Language:Chong, Source:Sir2001   4204 entries

IPA	GLOSS	LANGUAGE (DIALECT)	ID
tkɔːʔ	*worm*	Chong [of Kompong Som]	Isa2007b:C:2060-2
tɕʰiːʔm kəkʰeːŋ	*hornbill*	Chong [of Kompong Som]	Isa2007b:C:291-2
tɕʰiːʔm təkʰeːŋ	*hornbill*	Chong [of Kanchanaburi]	Isa2007a:C:291-1
(sa) ŋîː laː	*evening*	Chong [Kasong]	Nop2003:C:660-3
(sa) ŋîː troŋ	*noon*	Chong [Kasong]	Nop2003:C:1767-3
(sa) ŋîː troŋ	*noon*	Chong [of Ban Thung Saphan]	Huf1985:C:1767

multi-use dictionaries

Custom Dictionary
Language:Kensiu, Source:Bis1994   1635 entries

IPA	GLOSS	LANGUAGE (DIALECT)	ID
(ʔọʔ) bajuʔ nikah	*person being married*	Kensiu	Bis1994:C:1
-hah	*question marker*	Kensiu	Bis1994:C:2
-na-	*affix: nominalizer*	Kensiu	Bis1994:C:3
ahũn	*hiccup*	Kensiu	Bis1994:C:4
anɟedn	*to be complete, to be done, to be brown*	Kensiu	Bis1994:C:5

```
<customDict>
   <language>Kensiu</language><source>Bis1994</source>
</customDict>
<results>1635</results>

<item id="Bis1994:C:1" lang="Kensiu"><ipa type="phonetic">(ʔọʔ) bajuʔ nikah</ipa><gloss>person being married</g
<item id="Bis1994:C:2" lang="Kensiu"><ipa type="phonetic">-hah</ipa><gloss>question marker</gloss></item>
<item id="Bis1994:C:3" lang="Kensiu"><ipa type="phonetic">-na-</ipa><gloss>affix: nominalizer</gloss></item>
<item id="Bis1994:C:4" lang="Kensiu"><ipa type="phonetic">ahũn</ipa><gloss>hiccup</gloss></item>
<item id="Bis1994:C:5" lang="Kensiu"><ipa type="phonetic">anɟedn</ipa><gloss>to be complete, to be done, to be
<item id="Bis1994:C:6" lang="Kensiu"><ipa type="phonetic">aŋũt</ipa><gloss>throat</gloss></item>
<item id="Bis1994:C:8" lang="Kensiu"><ipa type="phonetic">bacuʔ bʌncɔʔ</ipa><gloss>clothing (compound)</gloss><
<item id="Bis1994:C:7" lang="Kensiu"><ipa type="phonetic">bacuʔ</ipa><gloss>shirt</gloss></item>
<item id="Bis1994:C:9" lang="Kensiu"><ipa type="phonetic">bagagŋ tu ŋoʎ</ipa><gloss>fungus</gloss></item>
```

```
Custom Dictionary
Language:Kensiu Source:Bis1994
1635 results
IPA       gloss     language (dialect)     ID
(ʔọʔ) bajuʔ nikah          person being married      Kensiu   Bis1994:C:1
-hah      question marker Kensiu   Bis1994:C:2
-na-      affix: nominalizer      Kensiu   Bis1994:C:3
ahũn      hiccup Kensiu   Bis1994:C:4
anɟedn    to be complete, to be done, to be brown Kensiu   Bis1994:C:5
aŋũt      throat Kensiu   Bis1994:C:6
bacuʔ bʌncɔʔ      clothing (compound)     Kensiu   Bis1994:C:8
```
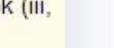
html
xml
text

dictionaries cooked…

(choose)
books by author
books by branch
all languages
branch+language

(sort)
○ IPA
○ orthography
○ gloss
○ ID

(display)
○ HTML
○ text
○ XML
○ JSON

Custom dictionary
Save file
Reset

University Press.
show all 229 entries :: build custom dictionary

Diffloth, Gérard
0000 **Unpublished field notes**. Cited in Diffloth 1980 "The Wa Languages".
show all 125 entries :: build custom dictionary

1976 **Notes**; cited in Diffloth 1980 "The Wa Languages,"
show all 73 entries :: build custom dictionary

1984 **The Dvaravati Old Mon languages and Nyah Kur** Monic Language Studies, Chulalongkorn University Printing House, Bangkok (iii, 402 pp., maps)
show all 3201 entries :: build custom dictionary

1980 **The Wa languages** *Linguistics of the Tibeto-Burman Area 5.2* (192 pp.)
show all 5837 entries / view source / cite :: build custom dictionary

1982 **Registres, dévoisement, timbres vocaliques: leur histoire en Katouique.** Mon-Khmer Studies 11:47-82.
show all 19 entries / view source / cite :: build custom dictionary

1989 **Proto-Austroasiatic Creaky Voice.** Mon-Khmer Studies 15:139-154.
show all 123 entries / view source / cite :: build custom dictionary

Drage, Godfrey
1907 **A few notes on Wa**. Rangoon, Superintendent, Government Printing.
show all 403 entries :: build custom dictionary

Returning complete dictionary for source "Dif1984"
Including both reconstructions and citations in search.
685 items found

MONIC BRANCH (685 ITEMS)

*( )pas   {to pick up // }   (proto Monic)   DIF1984:R:V254A
*[k] əpɔ̠h   {to pick up}   (proto Mon)   DIF1984:R:V254A.A
kəpɔ̠h   to pick up   (Mon [Rao])   DIF1984:C:V254A-2

* k-ləs   { // (spring-trap, arrow) to be released [Intransitive Verb]}   (proto Monic)   DIF1984:R:V260A
*khlə̠ỹ   {(spring-trap, arrow) to be released [Intransitive Verb]}   (proto Nyah Kur)   DIF1984:R:V260A.B
khlə̠h   (spring-trap, arrow) to be released [Intransitive Verb]   (Nyah Kur [Central])   DIF1984:C:V260A-1

*(-)ksaw   {red ant // red ant}   (proto Monic)   DIF1984:R:N41
*[k] əchaọ   {red ant}   (proto Mon)   DIF1984:R:N41.A
kəchaọ   red ant   (Mon [Rao])   DIF1984:C:N41-2
*ɕaw   {red ant}   (proto Nyah Kur)   DIF1984:R:N41.B
chaw   red ant   (Nyah Kur [Central])   DIF1984:C:N41-1

*-gɔh   {Noun Phrase particle // that (recently spoken of)}   (proto Monic)   DIF1984:R:N273
*-kɔ̠h   {Noun Phrase particle}   (proto Mon)   DIF1984:R:N273.A
kɔh   Noun Phrase particle   (Mon [Rao])   DIF1984:C:N273-2
*kɔ̠h   {that (recently spoken of)}   (proto Nyah Kur)   DIF1984:R:N273.B
khɔ̠h   that (recently spoken of)   (Nyah Kur [Central])   DIF1984:C:N273-1

*-m-raaŋ   {to open, splay, e.g. one's legs // open (flower) [Attributive]}   (proto Monic)   DIF1984:R:V80A

*-nɔɔʔ   {this, here // tonight}   (proto Monic)   DIF1984:R:N270

*-put   {to be cut off, truncated // (elongated object) to be truncated, cut in two [Intransitive Verb]}   (proto Monic)   DIF1984:R:V138

*-taaw   {to stand, station oneself, be positioned // (something) to be standing [Intransitive Verb]; (someone) to stand (something) up [Transitive Verb]}   (proto Monic)   DIF1984:R:V205A

*-truus   {boy, man // boy, man}   (proto Monic)   DIF1984:R:N180

*[ ]-ŋaak   {to look upward // to look upward (at something) [Transitive Verb]}   (proto Monic)   DIF1984:R:V24

*[ ]cɯɯt   {short time, moment // for a while, a while}   (proto Monic)   DIF1984:R:V132

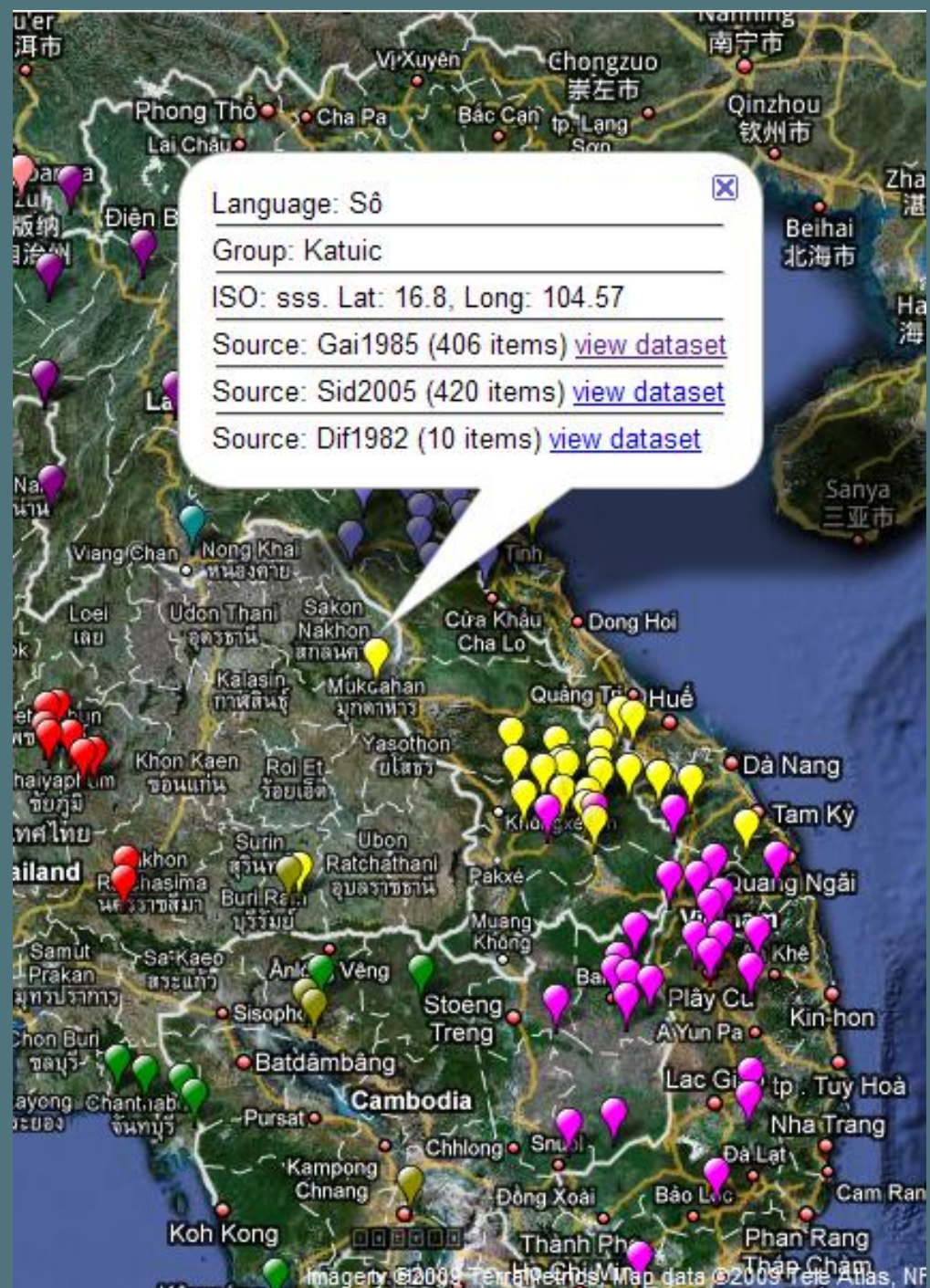*[ ]ləm   {to immerse oneself // to immerse something; to dip a fishhook}   (proto Monic)   DIF1984:R:V496

(choose)
books by author
books by branch
all languages
branch+language

(sort)
- IPA
- orthography
- gloss
- IP

(display)
- HTML
- text
- XML
- JSON

Custom dictionary
Save file
Reset

Huffman, Franklin
1985 **The Phonology of Chong**, in Suriya Ratanakul et al (eds.) *Southeast Asian Linguistic Studies Presented to Andre-Georges Haudricourt*. Bangkok, Mahidol University.
show all 544 entries / view source / cite :: build custom dictionary

1971 **Unpublished vocabulary lists**. 495.9, Mon-Khmer, Austronesian general folder, dated 1976. David Thomas Library, Bangkok, Thailand
show all 14032 entries / view source :: build custom dictionary

Isara Choosri
2007 **Investigating Contact-Induced Language Change: cases of Chung (Saoch) in Thailand and Cambodia**. PhD thesis, Mahidol University.
show all 551 entries :: build custom dictionary

2007 **Investigating Contact-Induced Language Change: cases of Chung (Saoch) in Thailand and Cambodia**. PhD thesis, Mahidol University.
show all 543 entries :: build custom dictionary

Kawa Pamphlets
1958-60 Various vernacular pamphlets published by the Yunnan National Publishing Company, cited in Diffloth 1980 "The Wa Languages".
show all 261 entries :: build custom dictionary

Kruspe, Nicole D.
2004 **A grammar of Semelai**, Cambridge, Cambridge University Press.
show all 1336 entries / view source :: build custom dictionary

Lefèvre-Pontalis, P.
1892 **Notes sur quelaues populations du Nord de l'Indochine**. Journal Asiatique (vol.& page nos. not specified)
show all 60 entries :: build custom dictionary

Fit Page

… and dictionaries raw

map accessible
dictionaries

maps by branch
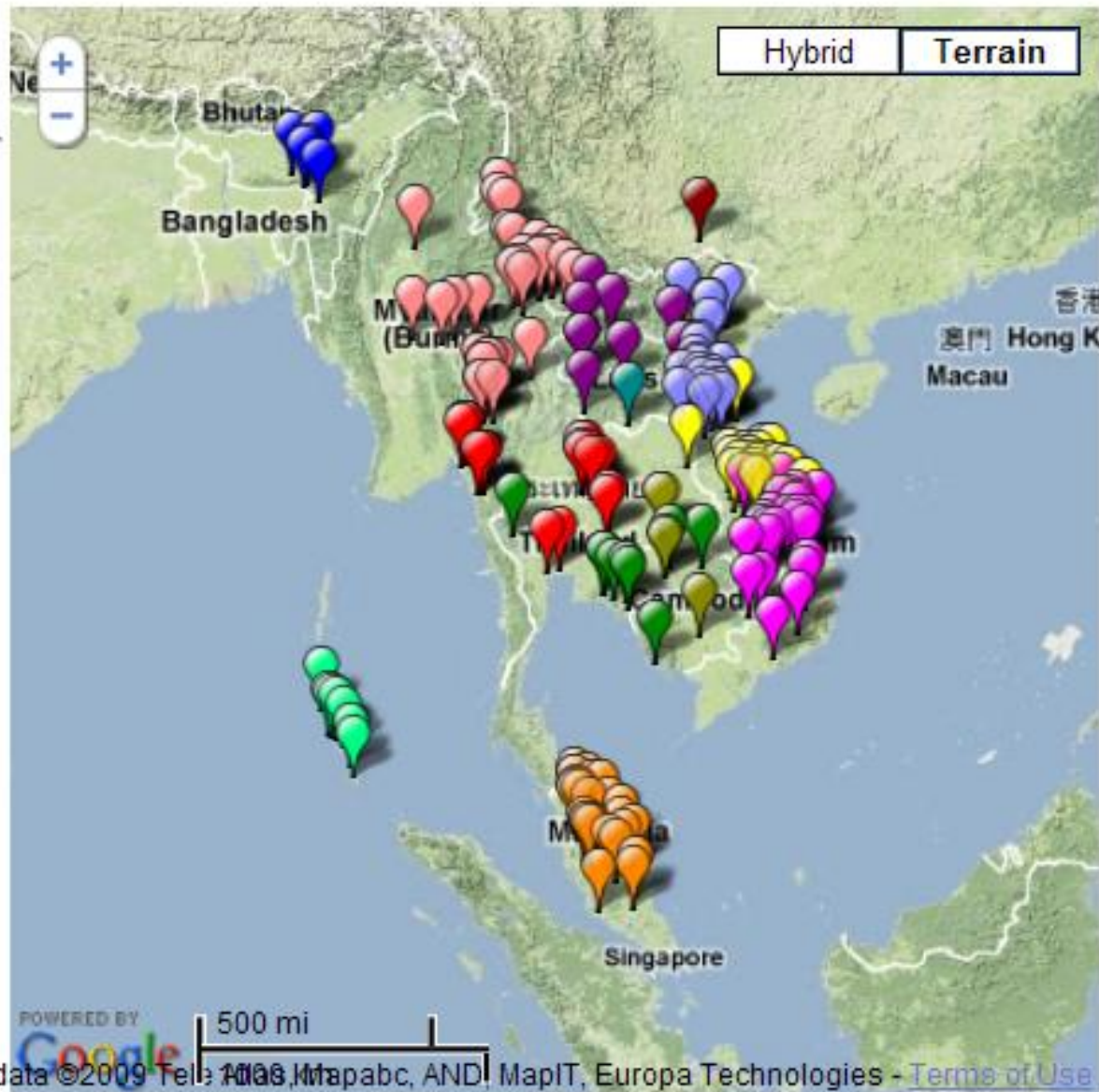
maps by count

Mon-Khmer Languages Project

Items per dictionary/source.
- ☑ 🟦 < 100 items (51)
- ☑ 🟥 100+ items (47)
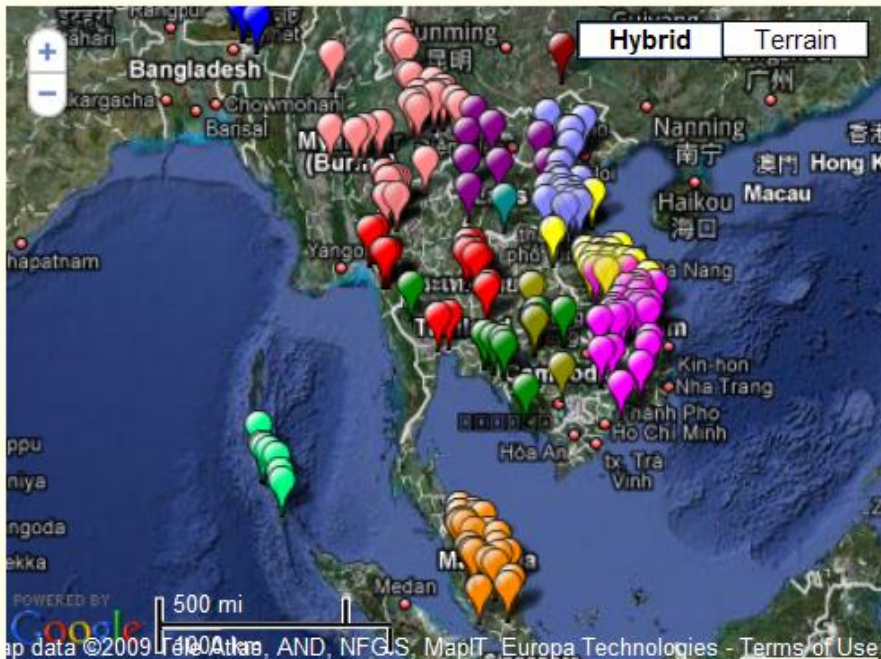- ☑ 🟩 500+ items (38)
- ☑ 🟨 1000+ items (50)
- ☑ ⬜ 5000+ items (10)

[On] [Off] [Reset]

view by counts / branches

maps for choosing

maps of the chosen

maps of what we found

Sunday, October 25, 2009 10:35:19 AM
IPA: [pbɓpʰ]V[nn̥tdd̥ʰθsʃll], **Text:** fly, **Return:** old to modern, **Sort by:** phonetic, **Search:** normal. **Language:** (*put mouse here*), **Include sources:** (*put mouse here*), **Exclude sources:** (*put mouse here*),

Sunday, October 25, 2009 10:33:45 AM
IPA: [pbɓpʰ]V[nn̥tdd̥ʰθsʃll], **Return:** full entry, **Sort by:** gloss, **Search:** normal. **Language:** (*put mouse here*), **Include sources:** (*put mouse here*), **Exclude sources:** (*put mouse here*),

Sunday, October 25, 2009 10:30:55 AM
Text: fox, **Return:** line-item, **Sort by:** phonetic, **Search:** normal. **Language:** (*put mouse here*), **Include sources:** (*put mouse here*), **Exclude sources:** (*put mouse here*),

Sunday, October 25, 2009 10:30:34 AM
Text: fox, **Return:** line-item, **Sort by:** phonetic, **Search:** normal. **Language:** (*put mouse here*), **Include sources:** (*put mouse here*), **Exclude sources:** (*put mouse here*),

Sunday, October 25, 2009 10:30:09 AM
Text: fox, **Return:** line-item, **Sort by:** phonetic, **Search:** normal. **Language:** (*put mouse here*), **Include sources:** (*put mouse here*), **Exclude sources:** (*put mouse here*),

and maps of where we looked

even a map of how to search

Please see the **MKED Tutorial and Cookbook** for more details.

In an **exact** search:
  B following a vowel requires breathiness.
  C following a vowel requires creakiness.
  D following a vowel requires a diphthong.
  L following a vowel requires a long vowel.

In a **normal** search:
- diacritics marking tone and phonation are ignored.
- long vowels and diphtongs are matched automatically.
- script and regular 'g' (g / g) are equivalent.
- ordinary 'h' matches IPA ^{h} for aspirated.
- IPA m̥ matches traditional **hm** for unvoiced.
- ':' matches IPA ':' for long vowels.

[abc]	allow any member of the set **a**, **b**, or **c**.
?	zero or 1 of the preceding.
[abc]? or **a?**	the set or letter is optional.
*	zero or more of the preceding.
.*	anything; zero or more letters.
[abc].* or **a.***	the set or letter followed by anything.
.*[abc] or .*a	the set or letter *preceded* by anything.
V	matches any vowel.
X	matches any consonant

In the past year, we have …

Cookbook/tutorial | Toggle branches | Toggle citations

Read 202105 citations and reconstructions, 99503 links, and 7886 notes from the database.
Built the etymological tree.
Too many hits (1837, limit 1500) to display except as line-items.
Showing 1500 items selected at random.
Building the full-entry view (this may take a moment).

much more data

a more efficient interface

Cookbook/tutorial   Toggle branches   Toggle citations

Original search target(s):
    gloss: "**house**" (as **house, houses, housing, housed**)
    language(s): "**212 languages.**"
Including both reconstructions and citations in search.
806 items found

MON-KHMER RECONSTRUCTIONS (94 ITEMS)

ASLIAN BRANCH (35 ITEMS)

BAHNARIC BRANCH (75 ITEMS)

KATUIC BRANCH (108 ITEMS)

KHASIC BRANCH (48 ITEMS)

KHMERIC BRANCH (56 ITEMS)

KHMUIC BRANCH (56 ITEMS)

MANGIC BRANCH (5 ITEMS)

MONIC BRANCH (123 ITEMS)

NICOBARIC BRANCH (56 ITEMS)

PALAUNGIC BRANCH (96 ITEMS)

PEARIC BRANCH (13 ITEMS)

VIETIC BRANCH (41 ITEMS)

MON-KHMER RECONSTRUCTIONS (26 ITEMS)

ASLIAN BRANCH (35 ITEMS)

BAHNARIC BRANCH (41 ITEMS)

KATUIC BRANCH (95 ITEMS)

x s *praaŋ *beam, joist* (proto Katuic) SID2005:R:132

x s *ʔabil *rat* (proto Katuic) SID2005:R:520

x s *til *fill a void/container* (proto Katuic) SID2005:R:525

x s *-peeŋ *behind* (proto Katuic) SID2005:R:663

x s *klɔɔm, *kalɔɔm *storage basket* (proto Katuic) SID2005:R:941

x s *-rɔh *burn(ing)* (proto Katuic) SID2005:R:1035

    rɔah *dry by fire* (Bru) MIL1976:C:SID2005~1035-1
    trɔh *spread (of a fire)* (Ta'Oi [of Sekong]) THE2001:C:SID2005~1035-2
    tarrɔh *burned (house)* (Ta'Oi [NVL]) NGU1986:C:SID2005~1035-3
    rɔh *spread (of a fire)* (Ngeq [Chatong]) THE2001:C:SID2005~1035-4
    ʔarɑh *spread (of a fire)* (Katu [Dakkang]) THE2001:C:SID2005~1035-5
    rɑh *on fire* (Katu [Triw]) THE2001:C:SID2005~1035-6
    rirɔh *spread (of a fire)* (Katu) THE2001:C:SID2005~1035-7
    rɑh *on fire, spread (of a fire)* (Kantu) THE2001:C:SID2005~1035-8
    ʔarɑh *burn* (Katu [An Diem]) COS1971:C:SID2005~1035-9
    rɑh *burning* (Katu [Phuong]) COS1971:C:SID2005~1035-10
    rɔh *burned (house)* (Pacoh) WAT1979:C:SID2005~1035-11

x s *tnool *post* (proto Katuic) SID2005:R:1100

untoggled branches and citations

extensive use of clustering

x s **ɲɨa?|haɛ?|hii?|ɲɲii|ɲii|həi...** *house... (cluster)* (proto Mon-Khmer) HUF1971:R:SID2(

x s **dùŋ|duŋ|dɨŋ|duŋ, doŋ|taŋ|toŋ...** *house... (cluster)* (proto Mon-Khmer) HUF1971:R:SID2009~3015

  dùŋ *house* (Kui) HUF1971:C:3015-408-8

  duŋ *house* (Souei) HUF1971:C:3015-408-9

  dɨŋ *house* (Bru) HUF1971:C:3015-408-10

  duŋ, doŋ *house* (Ngeq) HUF1971:C:3015-408-11

  taŋ *house* (Chong) HUF1971:C:3015-408-17

  toŋ *house* (Pear [of Kompong Thom]) HUF1971:C:3015-408-18

x s **hnəəm|hnaam...** *house... (cluster)* (proto Mon-Khmer) HUF1971:C:SID2009~3016

  hnəəm *house* (Laven) HUF1971:C:3016-408-13

  hnaam *house* (Brao) HUF1971:C:3016-408-14

x s **cəŋ|ciəŋ...** *house (classifier)... (cluster)* (proto Mon-Khmer) HUF1971:R:SID2009~3019

  cəŋ *house (classifier)* (Mon [Thailand]) HUF1971:C:3019-409-3

  ciəŋ *house (classifier)* (Thin) HUF1971:C:3019-409-6

x s **laŋ|hlaŋ...** *house (classifier)... (cluster)* (proto Mon-Khmer) HUF1971:R:SID2009~3020

  laŋ *house (classifier)* (Nyah Kur) HUF1971:C:3020-409-4

  laŋ *house (classifier)* (Thin) HUF1971:C:3020-409-5

  hlaŋ *house (classifier)* (Khmu) HUF1971:C:3020-409-7

  laŋ *house (classifier)* (Souei) HUF1971:C:3020-409-8

  laŋ *house (classifier)* (Bru) HUF1971:C:3020-409-9

building the citation trail

MIL1978:C:SID2005~1035-1
(Ta'Oi [of Sekong])   THE2001:C:SID2005~1035-2
(Ta'Oi [NVL])   NGU1986:C:SID2005~1035-3
Ngeq [Chatong])   THE2001:C:SID2005~1035-4
 (Katu [Dakkang])   THE2001:C:SID2005~1035-5
])   THE2001:C:SID2005~1035-6
 (Katu)   THE2001:C:SID2005~1035-7

fire)   (K   Theraphan L. Thongkum: 2001 ``Languages of the Tribes in Xekong Province Southern
Laos". The Thailand Research Fund. Bangkok, Chulalongkorn University.

Diem])   COS1971:C:SID2005~1035-9
ong])   COS1971:C:SID2005~1035-10
coh)   WAT1979:C:SID2005~1035-11

SID2005:R:1100

extracted dictionaries

Language: Katu

Group: Katuic

ISO: ktv. Lat: 15.12, Long: 108.38

Source: The2001 (1006 items) view dataset

Source: Sid2005 (2221 items) view dataset

SEAlang Mon-Khmer
Languages Project
Home / Dictionary / Database

SEARCH   Normal Exact
Reset

IPA

            Orth

house   Text (☑
expand)

Search: ☑ reconstructions ☑
citations

Return: Lay out / sort by:
○ old to modern   ○ ID

Custom Dictionary
Language:Katu, Source:The2001 1006 entries

GLOSS	IPA	LANGUAGE (DIALECT)	
a kind of palm similar to Caryota mitis (Palmae)	**tawaak**	Katu [Dakkang]	
abandon, throw away	**tuah**	Katu [Triw]	
able to	**bɤɤn**	Katu [Dakkang]	
above, upstream	**ʔapʌŋ**	Katu [Dakkang]	The2001:C:Sid2005~382-14
accuse	**toon**	Katu [Triw]	The2001:C:Sid2005~1120-7
ache, hurt, painful, sick	**kaʔaj**	Katu [Dakkang]	The2001:C:Sid2005~228-17

Choose languages | Choose IPA characters | Choose sources | Help | History | Issues | Map results

Language: Katu

Group: Katuic

ISO: ktv. Lat: 15.12, Long: 108.38

Source: The2001 (1006 items) view dataset

Source: Sid2005 (2221 items) view dataset

28	**Aslian**	pNorth Aslian	pCentral Aslian	pSouth Aslian	
27	**Bahnaric**	pN. Bahnaric	pC. Bahnaric	pW. Bahnaric	pCua
27	**Katuic**	pWest Katuic	pCentral Katuic	pKatu	
5	**Khasic**	pKhasic			
4	**Khmeric**	pKhmeric			
9	**Khmuic**	pPhray-Pram	pKhmu	pKhang-Khabit	
1	**Mangic**	pMangic			
20	**Monic**	pMon	pNyah Kur		
7	**Nicobaric**	pNicobaric			
33	**Palaungic**	pPalaung-Riang	pAngkuic	pLametic	
9	**Pearic**	pPearic	pChong		
25	**Vietic**	pThavung-Pong	pMương-Viet		

tools for deleting items and taking notes

| x | s | ***tnool** *post* (proto Katuic) S<small>ID</small>2005:R:1100 |

| x | s | ***ɟoon, *ɲon** *narrow (space)* (proto Katuic) S<small>ID</small>2005:R:1118 |

ɲǫn *narrow* (Ta'Oi [NVL]) N<small>GU</small>1986:C:S<small>ID</small>2005~1118-1

joun *narrow* (Ir) C<small>HI</small>1978:C:S<small>ID</small>2005~1118-2

coon *narrow (of space)* (Ngeq [Chatong]) T<small>HE</small>2001:C:S<small>ID</small>2005~1118-3

karcoon *narrow (of space)* (Katu [Dakkang]) T<small>HE</small>2001:C:S<small>ID</small>2005~1118-4

ɟoon *narrow (of space)* (Kantu) T<small>HE</small>2001:C:S<small>ID</small>2005~1118-5

con *narrow, crowded (land, houses)* (Katu [Phuong]) C<small>OS</small>1971:C:S<small>ID</small>2005~1118-6

ɲǫn *narrow* (Pacoh) W<small>AT</small>1979:C:S<small>ID</small>2005~1118-7

| x | s | ***puul** *shelter, cover* (proto Katuic) S<small>ID</small>2005:R:1235 |

| x | s | ***cuus** *shoo away* (proto Katuic) S<small>ID</small>2005:R:1271 |

Partial list on left (behind popup):

- X S *ʔabil *rat* (proto Katuic) SID2005:R:520
- X S *til *fill a void/co...*
- X S *-peeŋ *behind* (
- X S *kloom, *kaloom
- X S *-rɔh *burn(ing)*
- X S *tnool *post* (pr
- X S *joon, *ɲon *nar*
  - ɲo̰n *narrow* (
  - joun *narrow*
  - coon *narrow* (
  - karcoon *narr*
  - joon *narrow* (
  - con *narrow, c*
  - ɲo̰n *narrow* (
- X S *puul *shelter, co*
- X S *cuus *shoo away*
- X S ʔa.ʔu̯aj *a-oai* te
- X S ʔat *át* Imperfecti
- X S ʔa.viː *avi* not ha
- X S ɓu̯k *bóuc* bowl
- X S kaː *ca* relator th
- X S kəl.ləm tuk kən

Mon-Khmer Languages Project - Mozilla Firefox

http://nana/dictionary/popEty.htm

etyNotes.htm — Save as ...

Analysis of ..a few 'narrow' items — Comment / note.

save/print — Create a save/printer-friendly page.

*joon, *ɲon *narrow (space)* (proto Katuic) SID2005:R:1118
- ɲo̰n *narrow* (Ta'Oi [NVL]) NGU1986:C:SID2005~1118-1
- joun *narrow* (Ir) CHI1978:C:SID2005~1118-2
- coon *narrow (of space)* (Ngeq [Chatong]) THE2001:C:SID2005~1118-3
- karcoon *narrow (of space)* (Katu [Dakkang]) THE2001:C:SID2005~1118-4
- joon *narrow (of space)* (Kantu) THE2001:C:SID2005~1118-5
- con *narrow, crowded (land, houses)* (Katu [Phuong]) COS1971:C:SID2005~1118-6
- ɲo̰n *narrow* (Pacoh) WAT1979:C:SID2005~1118-7

We can save multiple items, with notes for each one.

kədriŋ... *narrow porch running all the way around ...* (CLUSTER) BAN1979:R:1227
- kədriŋ *koˀdrĭng* *narrow porch running all the way around house* (Bahnar [Pleiku]) BAN1979:C:1227-1
- kədriŋ *koˀdrĭng* *narrow porch running all the way around house* (Bahnar [Golar]) BAN1979:C:1227-2

More notes go here

Type notes for each entry as desired, then save items and notes by clicking the **save/print** button, above. Results from more than one MKL Project Dictionary search may be combined here.

saved items with notes

Downloads

Whew!

Part 3: Lessons

re-think **goals**

getting there is the problem

turn delivery into deliverables

re-think computing

We use computers as bookkeepers

We should use them as informants

and as ladders to the high ground

re-think **big science**

their big science Hubble**Cern**Genome

our big science STEDT**ACD**MKLP

re-think on-line

results + data
claims + tests
copies + sources

not just on-line libraries
on-line laboratories

re-think **sustainability**

it's not about maintaining income

it is about maintaining interest

it grows from

utility and accessibility

re-think language resources

ST|AN|KD|AA|HM
languages&structure

re-think language resources

Southeast Asian
linguisticsinfrastructure

Part 4: Questions

Search?
sealang.net / archives / seals2008.pdf

will it travel?

can this model be adapted to ST, AN, MK, TK?

will it code? do all daughters have mothers?

```xml
<item id="Dif1980:R:1" language="proto Mon">
    <ipa>*?u</ipa> <gloss>thingamajig</gloss>
</item>

<item id="Dif1980:C:1" language="Mon">
    <ipa>?ug</ipa><gloss>widget</gloss>
</item>

<link id="Dif1989:L:1"
    from="Dif1980:C:1" to="Dif1980:R:1" />

<note id="Dif1980:N:1" ref="Dif1980:L1">
  The plausibility of this relationship is
questionable.</note>
```
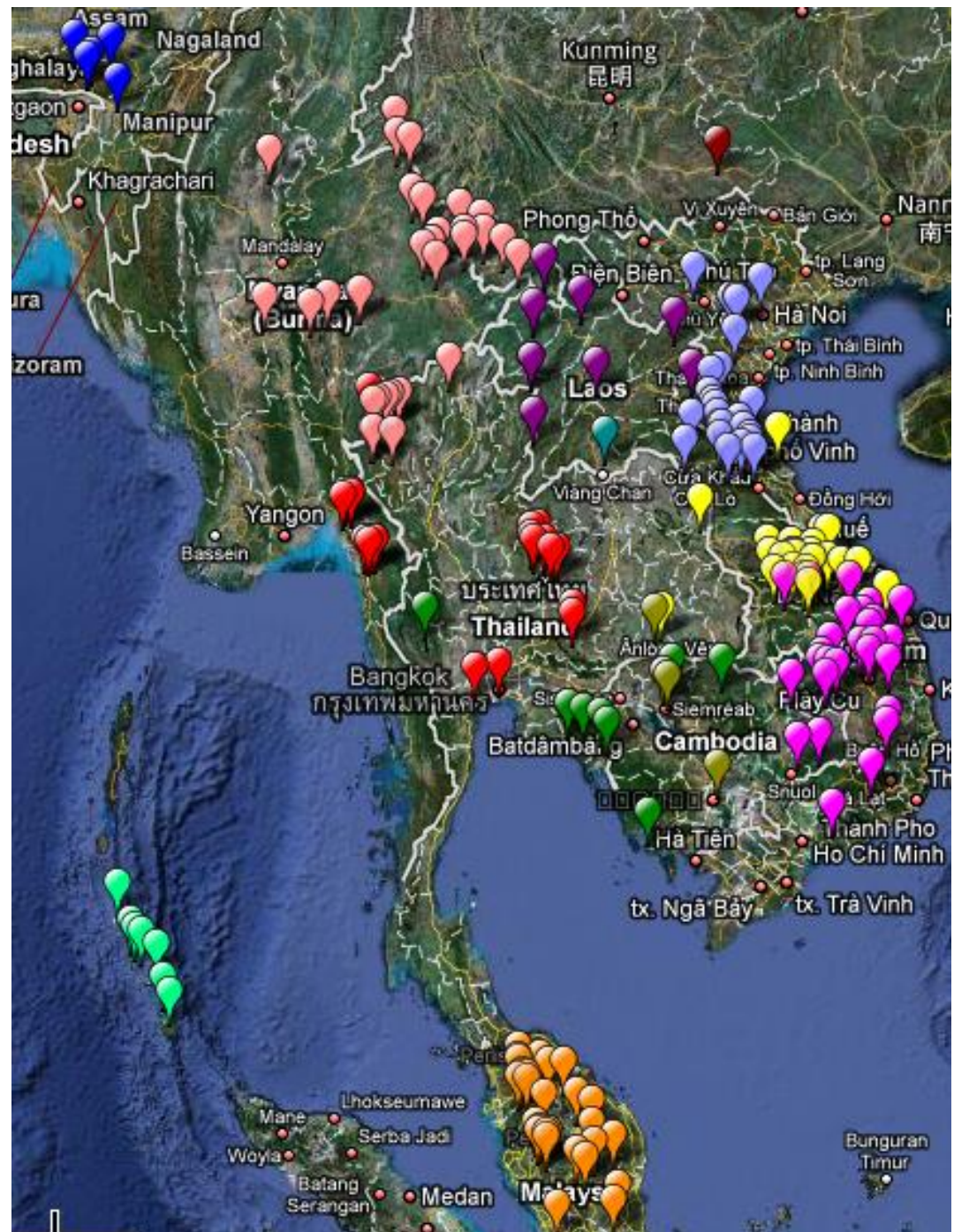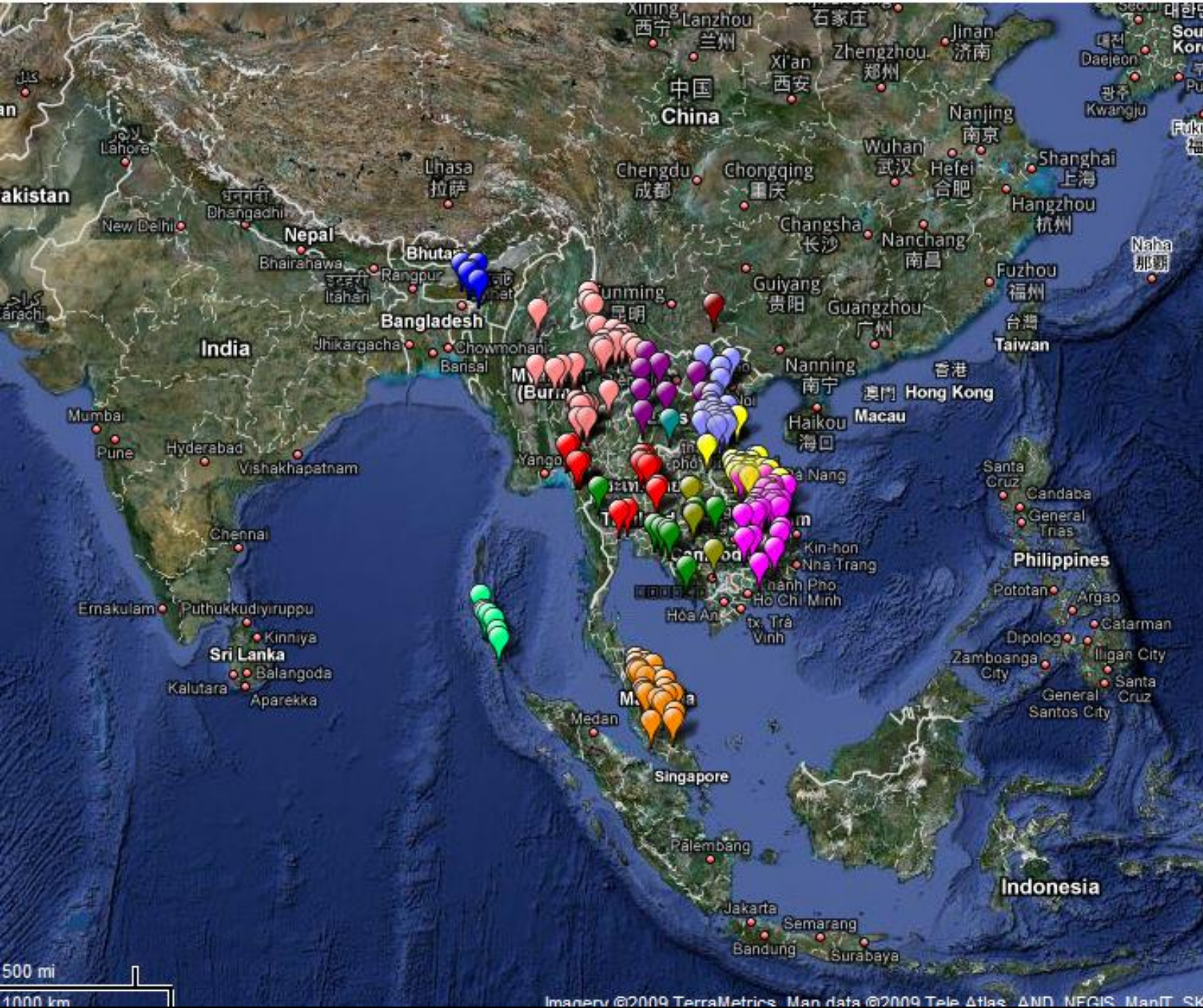
will it call?

http://sealang.net/api?
phone=.*g&
def="thingamajig"&
service=dictionary&
resource=mklp&
language=all&
sort=phonetic&
exclude=cooper2000&
extendAncestors=yes& ….

what's missing from this picture ?

the
part
that
isn't
there

Catch and release:
caring
feeding
sharing
of comparative data

Doug Cooper
Center for Research in Computational Linguistics

Mon-Khmer Languages Project
   National Endowment for the Humanities
   2007 – 2009, 2009 – 2011

http://sealang.net / monkhmer
http://sealang.net / archives / icstll42.pdf