Popularizing the Northern Khmer orthography: Sociolinguistics
in Action
Workshop on Northern Khmer orthography*
Surin, 23-24 April 1987. 26-27 May 1988. 22 March 1989.

Dorothy THOMAS, SIL and Mahidol University

1. Introduction
2. Background
3. Problems and Decisions

1. Introduction

There are upwards of a million people in northeastern
Thailand who speak a dialect of Khmer.   Various attempts at
writing it have been made by individual native speakers, and a
reasonable orthography using Thai letters had been worked out
for this Northern Khmer language[1] (henceforth N.K.) by William
Smalley and John Ellison.[2] This orthography has been used, with
some adaptations, for about 15 years.[3] A few things had been
published.[4] However, there are so many problems and varieties
of possible solutions, plus interference from the Cambodian
script for Standard Khmer, that there had been some objections
raised recently by influential northern Khmer people.

The hope was to get a consensus from the people themselves
and to get interest stirred up in writing this dialect of Khmer
with a minimum of conflict with the Thai writing system, yet
efficient for N.K. Many of the N.K. speaking villagers, although
somewhat fluent in spoken Thai, dropped out of school after
grade four, and are now only semi-literate. So such an
orthography would give the villagers a boost in reading ability
in general, as well as aiding in rural community development and
health programs and for the recording and preservation of their
culture.

So on April 23-24, 1987, May 26-27, 1988 and March 22,
1989, conferences were held in Surin Province, the heart of the
N.K. area in Thailand, to try to finalize an orthography. The
first two conferences were sponsored by the Institute of
Language and Culture for Rural Development of Mahidol University
and were chaired by Dr. Suwilai Premsrirat of the Institute. At
the first meeting there was a preponderance of delegates from
Surin, but the next year there was a more balanced
representation of Buriram and Srisaket provinces. Many
delegates were local officials who speak N.K. themselves, and
they were very interested and concerned, especially those from

the Department of Non-formal Education. The last conference
was held in conjunction with a writers' workshop, and was
jointly sponsored by the Mahidol Institute and the local office
of Non-Formal Education in Surin.

2. Background

Some of the problems of writing N.K. are accidents of
history, some are because of language differences, and some are
psycholinguistic. And of course these are intertwined.

The northern Khmer people have been cut off politically
from the southern Khmers of Cambodia for a long time and the
mountains form a natural barrier to frequent interaction with
the Cambodian Khmers. The approximately one million speakers of
N.K. are concentrated in Surin province, just north of the
Cambodian border, but there are also many in the provinces east
and west of Surin. The people see themselves as Thais who speak
Khmer.

With such a large number of speakers, the N.K. language
shows no signs of dying out. In Surin Province, the Khmer
speakers are even in the majority, and although in most of the
high schools the students are fined for speaking N.K., grade
school teachers have a real problem with the young Khmer
speaking pupils from the villages because they really don't know
Thai when they start school. There are many well educated Khmer
in high positions, including the (now) former provincial
governor of Surin, the mayor of Surin City, and many local
officials. The Khmer are not embarrassed to use their language
in many public situations, but for official communication they
use Thai.

Orthographically, the language with the strongest influence
on N.K. is Standard Thai. The area also has a lot of Lao and
Kuy villages, and some people speak all three languages, plus
Standard Thai. However, the Lao orthography is not well known
here so it has little influence on the writing of N.K. Kuy is
very non-prestigious, so it also has had little influence on
Khmer. Although N.K. and Cambodian are closely related, for
the most part, only men who learned to read Cambodian Buddhist
texts while in the monkhood can read the Cambodian script. So
the influence of the Cambodian script on the majority of the
northern Khmers is minimal.

Historically, the Thai borrowed much from the Khmer and Mon
languages and cultures. When the Thai became dominant in the
area, there was a lot of influence from Thai on Khmer so now
there are many borrowings and reborrowings in both directions,[5]

including Sanskrit and Pali words from the Buddhist literature. The ancient system of writing borrowed words seems to have been as in English: write the borrowed word as it is written in the donor language, not as it is pronounced in the receptor language. So now words borrowed at different times or from an intermediate donor are spelled differently, especially in Thai, although the meanings, and sometimes even the pronunciation, may be identical or closely related. Furthermore, although the Mon, Cambodian and Thai scripts are all derived from the Pallava script of India, each one developed differently. The languages of the region have developed a lot of extra letters as well as the irregular spellings. So one question is: how should N.K. spell these words?

Even a Thai child takes about as long to learn to read as his counterpart does in the English language. This creates an unbearable load for the beginning Khmer pupils. For a Khmer child to learn to read in his own language first would be a distinct advantage, but the likelihood that he would have a long period with Khmer classes in school is rather slim. So to have a fairly one-to-one correspondence between phonemes and graphemes in a N.K. orthography would be a big help for the children. However, some of the extra letters are conditioned in Thai, so although they may add an extra teaching load, they are not a big problem. And using them would aid in transfer between the two writing systems. So a completely one-to-one correspondence may not be necessary in N.K.

3. Problems and Decisions

3.1. Word level. the most surprising decisions of the conference were to write word spaces and where to position certain vowels. Both these decisions are very non-Thai, but the Khmer realized the ambiguities they would eliminate.

Thai writing has only minimal marking of word boundaries, making it sometimes difficult to identify word breaks. Unless the context is clear, you don't know if there is an open syllable, followed by an initial consonant cluster, or a closed syllable followed by a single (or double) consonant. Since N.K. has more possible consonant clusters and more complicated ones than Thai, the problem is multiplied without word breaks. The computer also needs word breaks for easy processing.

Certain vowels in Indian-derived scripts are written before the consonant that actually precedes them phonetically and phonemically. In some cases in Thai, and more often in N.K., this is complicated for two reasons. If there is a consonant cluster, Thai writes the vowel to the left of the whole cluster.

So it is not clear if there is one word with a consonant cluster or two words, one with the preposed vowel and one with the unwritten vowel. The Khmer decided to write word breaks, so this problem will not exist, but there is also a problem with simple three phoneme syllables with these preposed vowels. Are they CVC or CCV? There is at least one minimal pair in N.K. which really is ambiguous if vowels are written before clusters, but even for other words it will make reading simpler for a new reader to know at a glance if there is a consonant cluster or not. So it was decided to write these vowels before the consonant closest to the syllable nucleus.

3.2. Morphophonemically there are also a few problems, some similar to Thai, some not.

1) As in Thai, when there are two nouns in a noun phrase, if the first one has a long vowel, it is shortened. Thai writes these vowels consistently as long. This problem was not actually discussed, but undoubtedly the Khmer will want 'words spelled consistently'.

Thai treats these phrases as compound words: in citation form, no spaces between words; for line breaks, hyphenation between syllables. However, although the Khmer recognized some possible ambiguities, in 1989 they decided to write separate words.

2) N.K. has several functors which are unstressed open syllables, and the vowels are not clear. They are really clitics, but when written as an unstressed syllable of an adjoining word, they form an unusual word pattern. When written as a full word, they get too much stress. In Thai, similar functors can end with a final glottal when stressed, and are written with a glottal, although they frequently end in light aspiration or an open syllable. But this solution doesn't work for N.K. because /-ʔ/ and /-h/ are both in strong contrast with an open syllable.

At the 1988 conference it was decided to connect the clitics to the main word. But the 1989 conference reversed this decision, again showing the strong desire for breaks. As this conference was in conjunction with the writers' workshop, the delegates had had some experience in reading and writing following the previous decisions.

3) There is one functor /kəʔ/ which functions almost exactly like an analogous functor in Thai, but sounds a little different, especially when stressed. The Thai functor is written with no vowel but with the symbol for shortening a vowel

(ก๊). This is a particularly good solution for this functor in Thai because it looks so unusual, and it is a very common functor. For now they will try writing the N.K. functor as in Thai. People who see it for the first time usually read it as pronounced in Thai, but when they know it can be pronounced as in Khmer, they are satisfied. There is one problem for N.K. in that there is also a very common pronoun which has a somewhat similar shape (ก๊). Since this writing for the functor was adopted, some people have confused the two words.

4) Unlike Thai, sentence intonation is very important in N.K., usually affecting the last word of a clause. Thai does not use punctuation marks, and has only a break in the type to show sentence breaks. The conference was definite that they want more clause breaks written than in Thai.

There is a special intonation pattern for yes-no questions. It is either a lengthened final syllable or an added vowel. The added vowel does not form a true syllable as there is no initial consonant and it is unstressed. Instead of writing the question intonation-syllable, they will use the question mark.

3.3. Proper names are very frequently borrowed from Thai, Pali or Sanskrit. For personal names, people will probably want to write them as in Thai, but for geographical names they decided to write them first in N.K. style and then in Thai in parentheses. This will also be helpful when the N.K. name is actually different from the Thai name.

3.4. Literary loanwords will be written as pronounced in N.K., but followed by the Thai spelling in parentheses. Words like /thammada/ 'commonly', which have a double medial m will be written with the final am symbol +m-. (This particular word in Thai is written with a special symbol for th and a double r for the am, but would be written in N.K. as ท้ามดา [ธรรมดา]).

3.5. Northern Khmer sounds and symbols

The phonemic inventory of N.K. is as shown in Figure 1.

Consonants					Vowels			
ph	th	ch	kh	h*	i, ii	ɯ, ɯɯ	u, uu
p	t	c*	k		ɩ, ɩɩ	ɤ, ɤɤ	ʊ, ʊʊ
b	d				e, ee	ə, əə	o, oo
m	n	ɲ	ŋ		ɛ, ɛɛ	ʌ, ʌʌ	ɔ, ɔɔ
		r*				a, aa	ɒ, ɒɒ
		1*			iə, iiə	ɯə(ɯɯə)	uə, uuə
(f-in loans)	s						
ʍ	y						

Figure 1 Phonemes of Northern Khmer

Sounds not occurring in Thai are underlined. Consonants marked with an asterisk do not occur finally in Thai but do in N.K. /s/ is only initial in both languages. The length of the central off-glided vowel is predictable in Thai so it is written as a long vowel in Thai. However, in N.K. it is usually short, with the long vowel occuring in only one word.

1. Consonant symbols

	labial	alv.	palatal	velar	glottal
asp.stop	พ	ท	ช	ค	
unasp.stop	ป		จ	ก	อ-/-ะ,-ก
vd.stop	บ	ด			
nasal	ม	น	ญ	ง	
trill		ร			
lateral		ล			
fricative	(ฟ)		ซ		ฮ
semivowel	ว		ย		

Figure 2
1988 Northern Khmer Orthography - Consonants

One non-problem is what to do with the two sets of consonants in Thai which represent tonal classes. Everyone recognizes that N.K. has neither tones nor vowel register distinctions, so the set for mid tone is used. Another non-problem is the final voiceless stops: in Thai they are written as voiced stops. But in neither language is there voicing contrast in the final position so it doesn't matter. The real problems for the consonants are the phonemes that are only word-initial in Thai, and the one phoneme that doesn't

exist in Thai.

Although the palatal nasal does not occur in Central Thai at all, Thai does have a letter that represents /y/ initially and /n/ finally, and in some dialects of Thai it is actually pronounced as a palatal nasal. So it is a natural symbol for the N.K. palatal nasal. However, there is a problem because after front vowels in most dialects of N.K. there is no contrast beween the palatal and velar nasals. Except for Khmers who are very phonetically sophisticated, psychologically they are convinced that they are saying a velar. The tongue tip is clearly behind the lower teeth, causing the tongue to hump in the middle, and the phonetics professor assured the delegates that they were actually saying a palatal, but they remained unconvinced.   At the first conference they side-stepped the issue, but at the 1988 conference they decided to write the velar symbol following front vowels. The only problem that concerned those who had finally realized how they pronounce the consonant was whether writing a velar for what is a really palatal would confuse their children. During the following year there was a strong reaction against writing the velar for the palatal.   So in 1989 they decided that they would write the palatal, at least for widespread printing, and locally people could do as they liked.

The final gutturals cause a problem because in Thai the final glottal stop sometimes is realized as [h], and in N.K. the final glottal stop is sometimes realized as [k]. Thai has no phonemic final /h/, but N.K. does. Many dialects of N.K. have no phonemic final /k/.

However, the k-ʔ contrast may be coming in from Thai, and there is a strong preference for writing /k/ after the short central and back high vowels. Thai has no glottal stops after long vowels, but N.K. does. So in the Smalley-Ellison orthography the Thai symbol for /k/ is used after long vowels. Probably they will continue using the /k/ symbol after the long vowels. After the short high central and back vowels, although they usually say a glottal, many Khmer are convinced that they are saying a velar. So the conference decided to write the velar instead of a glottal after the short high central and back vowels, although the Khmers who are used to writing in the existing orthography do balk at the arbitrary spelling.

The final /h/ came in for much discussion because it is so non-Thai. A few Thai words are written as if they had a silent /h/, but a different symbol is used from the one used initially for words with level tones. To complicate matters, Cambodian uses an /s/ symbol for final /h/. But eventually the symbol

used for initial /h/ in N.K. was adopted.

Surprisingly, there was little discussion of the final /r, l, c/. Although Thai does not have these finals, there are comparable symbols used finally in Thai, but they are pronounced as /n, n, t/. Some minimal pairs were presented and everyone seemed satisfied to write the phonemes as pronounced in N.K. The Khmers know that the Thai pronounce these sounds wrong when they try to speak N.K.

Consonant clusters are not written consistently in Thai. They decided to write them consistently, regardless of whether there is a slight open transition between the consonants or not.

Syllabic nasals are not very common in N.K., but they do occur before consonants in some words. The 'obvious' solution of just writing a simple nasal is not acceptable because it would be read with an intervening (unwritten) vowel. The Smalley-Ellison solution of writing as if there were a presyllable /ʔaN/ was adopted, although it does take some learning.

2. Vowel symbols.

The symbols chosen for the vowels are shown in Figure 3.

	short ʔVd ʔVʔ	long Vd	short ʔVd ʔVʔ	long ʔVd V#	short ʔVd ʔVʔ	long Vd
high	อิค/อิ	อีค	อีค/อิก	อีค/อือ#	อุค/อุก	อูค
high open	อฺิค/อฺิ	อฺีค	อฺีค/อฺิก	อฺีค/อฺือ#	อฺุค/อฺุก	อฺูค
mid	เอ็ค/เอะ	เอค	เอิค/เออะ	เอิค/เออ#	อฺือค/โอะ	โอด
mid open	แอ็ค/แอะ	แอค	แอฺิค/เออะ	เอฺิค/เออ#	อฺือค/อะ	ออค
low	(อัค, เอา, อำ, ไอ)/อะ*		อาค/อา#		อฺือฺุค/ออะ	ออูค
glided	เอียฺ็ค	เอียค	เอีอฺ็ค	(เอือค)	อฺัวค/อัว#/อัวะ	อัวค

−d = any consonant but glottal
*The varient symbols for short /a/ represent:
/ʔad, ʔaw, ʔam, ʔay and ʔaʔ/

Figure 3
1989 Northern Khmer Orthography - Vowels

As an aid to visualizing the positions of the vowels, the examples show an initial glottal and either a final d or a final glottal. (Note that the symbol for initial glottal is the same as for parts of the mid-central vowels and the low back vowels.) Final glottal is unwritten after some short vowels. For open syllables (indicated here with #), some long central vowels add the symbol for initial glottal/low back vowel after the vowel. The shortener (ˇ) has to be used on more vowels than in Thai.

The one vowel that was radically changed was the open mid-central vowel. Formerly it was written แ◌ˇ, as the central counterpart of the open mid front vowel (building another symbol onto the front symbol, as is already done for the close vowel as in Thai), but it looks strange in loanwords. So they chose to write เ◌ˇ, following the pattern for lowered vowels (same symbols as the close vowels, but with the added dot beneath).

The special symbols for short /a/ before /m/, /w/, /y/ and /ʔ/ will be used as in Thai (◌ัว, เ◌า, ไ◌, ◌ะ). This is especially important for /aw/ because the Thai symbol for /uu/ uses the normal expected symbols for /aw/ (◌ัว). A problem may arise with the /am/ and /ay/ because in loanwords from Sanskrit-Pali, Thai writing is not consistent, and some of those very same words are common in N.K.

The central off-glided vowel is very short in N.K., but long in Thai. Time will tell whether people will consistently write the symbol for vowel shortening or not.

For the lowered vowels, the present system was accepted: a dot below the line, except when the vowel itself is already written below the line. Then a circle above the vowel is used. The only thing this changed from the previous orthography is that it had the raised circle instead of the lowered dot for the only lowered vowel which is written on the line (-◌).

Thai usually writes the vowel shortener above the initial consonant, but this is complicated if there is already a mark there. And for the computer, this is a disaster. However, there is some precedence from Thai to shift the shortener to the right (for newspaper headlines), and since we began doing this several years ago, there has been little reaction one way or the other from the Khmer readers. Probably for handwriting, people will put the shortener on top of other diacritics, but they accepted the shifted position for printed materials.

The 1989 conference addressed the problem of limited distribution of certain short vowels in some environments. Some villages tend to one pronunciation, others to another, and some

villages may have contrast. While allowing leeway for local
situations, they decided that for widespread printing one should
follow these conventions:

1)   [ʋ] and [o] before bilabials:

Use [ʋ] for ยุ๋บ 'night', กรุ๋บ 'all', จำปุ๋บ 'stumble'

ยุ๋ม 'cry',   พนุ๋ม 'mountain'

[o] (the unwritten vowel)

for ดบ 'overgrown', กนบ 'grasshopper'

ทม large', ซม 'request',   คลม 'blow',

อม 'aunt, uncle'

2)   [ɪ] and [e] before palatals

Use [ɪ] for หมิ๋ญ 'tooth', ปิ๋ญ 'full'

พลิ๋จ 'forget', กิ๋จ 'step away'

[e] for เดิ๋ญ 'know'

3)   [e] and [ɪ] before /y/ use [ɪ] as in ปริ๋ย 'jungle'

4)   [e] and [ɛ] before /y/ use [ɛ] as in ซแร๋ย 'woman'

In 1989 they also decided how to write short vowels before
glottal where the writing system is open to two possibilities
and Thai uses an unusual symbol or digraph + glottal.   Rather
than using the shortener (˘) plus /k/, they will use the special
Thai symbol for the short vowel + final /ʌ, ɒ/ (เ-อะ, เ-าะ)
following the Thai pattern for the next higher vowels /ə, ɔ/
(เ-อะ, เ-าะ): เพอะ 'drink' เซาะ 'hair' As in Thai, they will
write a long /oo/ instead of using the short, unwritten /o/
before glottal (โ-ะ): โปะ 'to filter'.

NOTES

*An earlier report on the first Northern Khmer conference was
published in <u>Notes on Literacy</u>, [Dallas, SIL.] 1989, Vol.57,
pp.47-57.

1.   See David Thomas, On the "language" status of Northern
Khmer.   Paper presented at the Sino-Tibetan Conference, Lund,
October 1988, and to appear in the Mahidol <u>Journal of Language</u>

and Culture.

2.   See William Smalley. 1976.   The problem of vowels:
Northern Khmer.   In Phonemes and Orthography: Language Planning
in Ten Minority Languages of Thailand. William A. Smalley, ed.
Pacific Linguistics, Series C, No.43. pp.43-83.

3. See Dhanan Chantrupanth and Chartchai Phromjakgarin.   1978.
Khmer (Surin)-Thai-English Dictionary. Indigenous Languages of
Thailand Research Project, Chulalongkorn University Language
Institute, Bangkok.

4. As Kheuan Singkhanipa, Suwilai Premsrirat and David Thomas.
1984.   Surin Khmer Conversation Lessons. Mahidol U., Bangkok.
And a four language vocabulary with Thai, Kuy, Northern Khmer
and English.   (by Feikje van der Haak and Brigitte Woykos).
1986. Summer Institute of Linguistics and Institute of Language
and Culture for Rural Development, Mahidol U. Bangkok.   (No
English title is given.) Plus various scripture portions.

5.   See Franklin E. Huffman. 1986. Khmer loanwords in Thai.
In Papers from a Conference on Thai Studies in Honor of William
J. Gedney. Robert J. Bickner, Thomas J. Hudak and Patcharin
Peyasantiwong, eds.   Michigan Papers on South and South East
Asia, no.25. University of Michigan, Ann Arbor, MI, pp.199-209.

22 June 1989   Institute of Language and
                                    Culture for Rural Development

                            *
                    *   *