

NOTES ON KRAISRI'S AND BERNATZIK'S WORD LISTS

William A. Smalley

The Director of the Siam Society Research Center has very kindly invited me to study the pre-publication manuscripts of the word lists of *Mrabri*, *Yumbri*, *Khmú*?, and other languages listed for comparison in this special issue of the Journal. I offer the following notes and observations in appreciation of the effort which has gone into the publication of this very valuable linguistic material, and in the hope that they will enhance the usefulness of these lists through the interpretations they offer.

The lists present several fundamental challenges to study which I have not had opportunity to follow up. For one thing, it would be useful to see what kinds of phonemic systems these languages present, in the light of the amount of data included. I have not had access to the tape recordings which Mr. Kraisi collected for *Mrabri*, however, and an analysis of the sound system should wait for that.

A detailed comparative study of the systematic correspondences between these languages would also be in order, since so much of the data is new. This would entail a word-by-word, or rather phoneme-by-phoneme analysis of the regularities of correspondence, and the postulation of the forms from which present speech is derived. It would also involve placing this information in relation to the wider linguistic groupings of this language family. Such an analysis will have to wait more time.

Two kinds of problems were amenable to study in the short time available, however, and the results of this study follow in this paper. I made a brief statistical comparison of basic vocabulary between *Mrabri* and the other languages, and an analysis of the transcription systems used in the one case by Kraisi for *Khmú*?, and in the other case by Bernatzik for *Yumbri*. In so doing some additional light is shed on the lists, making them more useful to others.

Comparisons of *Mrabri* Vocabulary

Mr. Kraisi has done a great service in collecting and compiling a comparison of *Mrabri* vocabulary with Bernatzik's *Yumbri* and with the

three dialects of Tin, with Klmu² of Muong Sai, and with dialects of Khamed, Lawa, and Mon. The lists make fascinating study, and will be a contribution to the data needed for comparative analysis of linguistic relationships.

Ultimately questions of linguistic relationship must be solved by painstaking comparative analysis, word by word, between two languages or within groups of languages. In such work linguists look for more than "similarities". They look for regularities of correspondence, where the same phoneme in one of the languages regularly corresponds to a phoneme in the other language under the rules which can be described. In light of our limitations here we will have to do something more provisional, less fully diagnostic, based on similarities of "basic vocabulary" only.

Linguists find that for purposes of gaining a quick impression of linguistic relationships not all words are equally useful. Some linguists speak of "basic vocabulary" in referring to a small number of words not very likely to be borrowed from one language to another because they are all so basic to human experience. Such words include the words for 'man,' 'women,' 'person,' 'fish,' 'dog,' 'come,' 'cat,' 'sun,' etc.

One such list, often called the "Swadesh list" after its compiler, consists of two hundred words in its longer form. I have made a quick analysis of Kraisi's and Bernatzik's material as found on this list, and a comparison of Mrabri with each of the other languages in the lights of this selection of words.

Of the two hundred items on the Swadesh list, Kraisi lists sixty-six and Bernatzik fifty-eight. I compared these "basic vocabulary" items in Mrabri with each of the following languages, and arrived at the number of obviously related words (first column), percentage of relationship (out of the 66 or 58, as the case may be, second column). I also included Kraisi's own figures for his whole lists (not just basic vocabulary, third column).

Language	No. of Related Words (Basic)	Percentage of Relationship	Total Related Words (Kraisri)
Khmu ²	25	38	35
Yumbri	20	35	41
Tin Namyoh	23	35	} 45
Tin Sagad	22	33	
Tin Ben 'Toci	21	32	} 20
Khamed	13	20	
Mon	12	18	13
Lawa	10	15	14

Related Basic Vocabulary Between Mrabri² and Other Languages

I feel that this scanty evidence does not support Mr. Kraisri's contention that the Mrabri and the Yumbri are the same people. According to these figures the Mrabri are linguistically just as closely related to the Khmu² and the Tin as to the Yumbri. In the light of the evidence I doubt very much that Mrabri and Yumbri would be mutually intelligible. That Bernatzik's data are faulty is very possible, but the evidence as it stands would make Mrabri just as different from Yumbri as it is from Khmu² and Tin. These latter are certainly not "the same people". That these languages are in some way related to Mon-Khmer I think is incontestible, but it might be wisest for now simply to say that they are part of Austroasian (of which Mon and Khmer are also a part) without tying them more closely to Mon-Khmer than to other branches.

Kraisri's Transcription of Khmu²

Because I have some knowledge of Khmu² and have prepared a descriptive grammar¹ of it I have studied Mr. Kraisri's previous transcription of Khmu² in the *Journal of the Siam Society*² and in his present paper in this issue, where in both cases he lists Mrabri words with Khmu²

1. William A. Smalley, *Outline of Khmu Structure*, Essay No. 2 of the American Oriental Series. New Haven, Conn. American Oriental Society, 1961.

2. Kraisri Nimmanahaeminda and Julian Hartland-Swann, "Expedition to the "Khon Pa," JSS Vol. L Part 2 (July 1962), pp. 180-183.

equivalents. Judging by Khmu², Kraisi's remarks on his Thai transcription as against his romanized one are well taken. For someone untrained in phonetics and the adaptation of roman symbols to phonetic use, the Thai script lends itself to transcribing Khmu² better than do English writing habits.³ From a linguistic standpoint Mr. Kraisi's Thai transcription of Khmu² is fully workable with some exceptions.

The Muong Sai dialect of Khmu² which Mr. Kraisi recorded is slightly different from the Luang Prabang dialect with which I am most familiar. I will not comment on differences which are obviously due to dialect. There are some purely transcriptional problems, however.

In the writing of any language certain conventions have to be established. It is almost always impossible to take over the writing system of one language for use in another and have it represent the phonemes (significant sound distinctions) adequately. A phonetic or phonemic transcription in romanized script uses the symbols in ways accepted by convention among schools of linguists, explaining any deviation from that convention. However, taking over English writing habits as such for the transcription of another language is extremely difficult because of the inconsistency in English spelling.

Thai transcription is potentially more consistent, and its inventory of phonemes happens to be quite similar to those of Khmu². But the Thai writing system likewise has to be adapted for Khmu². Some Thai symbols have to be used with new values if all Khmu² phonemes are to be recognized. Some new combinations of consonants are going to have to occur with new values, and the tonal system of Thai will have to be discarded. In making such an adaptation the new conventions need to be made explicit for the benefit of the Thai-reading person, and of the linguist who might try to figure out the phonemes by means of the Thai. It is for a clarification of some of these points that the following discussion is presented.

3. An extensive paper which I have written on the problems of writing Meo, Yao, Kuy, Lawa, and Karen in Thai script will appear in William A. Smalley, and others, *Orthography Studies*, to be published (1963?) by the United Bible Societies, London.

Mr. Kraisri uses some Thai tone indications in transcribing Khmu² and the other languages he lists. Khmu² (that is any dialect I have studied) is not a tonal language. If the others are, it is to be proven.⁴ When Thai consonants are used they must be used without any sense of the tone classes they represent in Thai. *๑* and *๒* would mean the same thing in Khmu². There is no point in using both.

On Khmu² final consonant which are pronounced why does Mr. Kraisri write *๑* or *๑*? For example in his *๑๑๑๑* /hi^๑iar/๑ why this mark on the final *๑* /r/? It is enough to state that all consonant are pronounced as written, whether initial or final.

Khmu² has a contrast (at least in the Luang Prabang dialect) between /c/ *๑*, /ch/ *๒*, and /j/ for which there is no symbol in Thai. (Mr. Kraisri romanizes *๑* as j, which is satisfactory if he also has a symbol for what we call /j/.) I cannot be positive that this contrast between /c/ and /j/ exists in the Muong Sai dialect, but think it likely. It is not represented in Mr. Kraisri's transcription. In Luang Prabang, for example, /c^๑iaŋ/ is the name of a legendary culture hero, /j^๑iaŋ/ the word for 'foot'. I think these would both be written *๑* in Mr. Kraisri's script.

Khmu² has /c/ *๑* in final position, pronounced as such. People not familiar with it, often miss it however. Mr. Kraisri transcribes it in various ways. In the previous JSS article he transcribes /hmu^๑c/ 'ant' as *๑๑๑๑*. In this new material he makes it *๑๑๑๑* which is phonetically closer in the final consonant. Actually, by the structure of the Khmu² language, however, it should be *๑๑๑๑* (on *๑* see the next paragraph). The *๑* /y/ sound is caused by the presence of the /c/ and should not be written because it is automatic. Kraisri transcribes *๑๑๑๑* 'hear' for *๑๑๑๑* /mec/.

4. It is not true to say, as some have done, that Mon-Khmer languages are not tonal. Some are. In fact, David Filbeck informs me that there is strong evidence for a rudimentary tone system in Tin. However, it is unlikely that most of the languages listed by Kraisri are tonal.

5. Transcriptions in / . . / are my phonemic transcriptions. /ʔ/ is the symbol for glottal stop, or catch in the throat which serves as a consonant in Khmu² and many other languages.

The previous paragraph contains another set of problems. The Khmu[?] I know has /hw by hn hn hñ hŋ/. The hñ in these cases represents that these sounds are made without any vibration in the vocal cords (voiceless, or sourd). The mouth is set the same as for /w y m n ñ ŋ/, but the sound is very different. In Luang Prabang Khmu[?], for example, /ŋw[?]/ ၂၅၃ is 'fear' but /hŋw[?]/ ၂၅၃ is 'paddy'. Mr. Kraisi transcribes 'tooth' as ၂၅၃ (i.e., /raaŋ/) whereas in Luang Prabang it would be ၂၅၃ /hraaŋ/. Again Thai readers should be warned that the ၂ is used in this combination by a convention which does not occur in Thai. It is used here for a consonant sound, not as an indication of tone. Quite likely at an earlier stage of Thai development ၂၅, ၂၆, etc., in Thai represented the sound we are now using it for in Khmu[?], and developed its tonal association when the sound became voiced in the historical development of Thai.

I welcome Mr. Kraisi's very valuable Khmu[?] lists. These are the best lists of Khmu[?] vocabulary to date.⁶ They will aid comparative studies. Perhaps these notes will help the linguists who use them to interpret them somewhat for linguistic uses.

Bernatzik's Transcription of Yumbri

The republication of Bernatzik's little-known word list of Yumbri would be a great deal more valuable if it could be "matched" linguistically with Kraisi's transcription of Mrabri. That is, if the same system of transcription could be used for both, and if we could be sure that the premises and conventions followed in the two transcriptions were somewhat alike it would make the parallel listings much more valuable. Since I have some understanding of Kraisi's transcription of Mrabri, assuming he followed the same conventions he used in recording Khmu[?], I thought it might be worthwhile to make an analysis of Bernatzik's transcription of Yumbri to see where it might be clarified for purposes of comparison.

6. The only other list is in Henri Roux and Tran-van-Chu, "Les Tsa Khmu," in *Quelques Minorités Ethniques du Nord-Indochine*, Saigon: *France Asie* Nos. 92-93, pp. 297-357.

My grammar contains nearly seven hundred different Khmu[?] word used in the various illustrations, but these are not in any list form. Rather, they are sprinkled through the pages of the book.

The method used was that of a brief structural analysis of some of the most obvious features of Bernatzik's transcription in the light of experience in the analysis of several related languages. The analysis is not exhaustive, and I do not pretend that it is "phonemic," but I think it gives us a basis for a retranscription of much of Bernatzik's material in a more useful form.

The consonants are easier to handle than the vowels, so we will take them first. To simplify the problem even more, we will take the consonants which occur at the end of Yumbri words as listed by Bernatzik before we go on to others. Rather than listing these alphabetically (which has no linguistic meaning) we will list them according to their presumed phonetic qualities. Here we assume that Bernatzik used them in a way conventional to European languages he knew or to phonetic usage. In the following chart the first column contains sounds made by closing the lips (bilabial), the second by closing the tongue behind the teeth, the third a little farther back in the mouth, and the fourth in the back of the mouth.

b			g
p	t		k
m	n		ŋ
	l		
	r		

y

Bernatzik's Final Consonant Symbols

Linguistically, the placement of each of the above symbols is fairly easy. In the case of *ŋ* we are aided by Bernatzik's own explanation.⁷

Once we are satisfied that the symbols are correctly placed, as nearly as we can tell according to their likely phonetic properties, we look at the chart to see what linguistic problems it presents. In the first place we are struck by two pairs of symbols *b/p* and *g/k*. In this part of the world I know of no language where the sound [b] contrasts with the sound [p], or the sound [g] with the sound [k] at the end of a syllable. All of our symbols were taken from the ends of words, so there is probably

7. Hugo Adolf Bernatzik, *Die Geister der Gelben Blätter*, München: Verlag F. Bruckmann, 1938, p. 237,

no significant difference within these pairs. Here we are referring to pronunciation. In Thai, for example, the characters *๑*, *๒* may both be used as the final *letter* in a word, but they both carry the same *sound*. There is no spoken contrast in that position. [b] and [p] make a difference in many languages of this area when they occur at the beginning of a syllable but not at the end. Our first reaction, on the linguistic level, is to consider the writing of *b* and *g* at the end of the syllable superfluous. We write *p* and *k* instead because in most of these languages the sounds in question have more of the phonetic qualities of [p] and [k] than of [b] and [d].

The second column we notice has to do with column three in the chart. Languages of this family usually have some more sounds in this position. When we take another look at Bernatzik's material, to include his clusters of two consonants at the end of his words we find the following: *dy*, *ty*, *yn*, *gn*. We insert them in the chart as follows.

		dy	
p/b	t	ty	k/g
m	n	gn	ñ
		yn	
	l		
	r		
		y	

Bernatzik's Final Consonant and Consonant Cluster Symbols

We immediately eliminate *dy* in favor of *ty* on the same basis as *b/p* above. We assume that the *ty* is the same sound /c/ as is transcribed *๑* in Thai (and in Khmu², as described above), but that unlike Thai (but like Khmu³) it occurs at the end of a syllable in Yumbri. We are of course referring to pronunciation, not just to writing. It is very common at the end of the syllable in Cambodian and other Mon-Khmer languages.

We assign *gn* and *yn* to a sound [ɲ] as in Spanish. There is probably no distinction between *gn* and *yn*. The sound occurs at the beginning of a syllable in Northern Thai and in Lao. It occurred at an earlier stage of Bangkok Thai and was written *๑*. We assume that Bernatzik meant this sound by *gn* because that is the way it is spelled in

French. As for *yn*, we are assuming that Bernatzik made a very common mistake in transcription. In many of the Mon-Khmer languages, including Khmer itself and Khmu², when the [ɲ] occurs finally after certain vowels there is a little automatic [y] sound before it, and the unwary transcribe it as *yn* (cf. the discussion of Khmu² /c/ which is a parallel problem). We therefore retranscribe Bernatzik's *gn* and *yn* the same way because they probably represent the same phoneme.

Another problem remains in the chart. There is no *w*, which we would expect in this area. We find that Bernatzik has used three vowel symbols, *u*, *û*, and *ǔ* in final position after vowels. For example, in the word for 'chicken' *ǎtǔû* we assume that the *û* represents /w/ because that is what /w/ would sound like after this vowel. Our further reasoning for this analysis here is a little technical in nature, and we will not go into it. The chart now stands as follows.

p/b	t	ty/dy	k/g
m	n	gn/yn	ñ
	l		
	r		
u/ǔ/ǔ		y	

Reinterpretation of Bernatzik's Final Consonants

The resulting chart is beginning to look like what we would expect of languages in this area. Compare the corresponding chart in Khmu²:

p	t	c	k	ʔ
m	n	ñ	ŋ	
				h
	l			
	r			
w		y		

Khmu² Final Consonant

Not all Mon-Khmer languages are identical, of course, in their charts of final consonants, and we should not expect them to be, but many are similar. Two differences emerge in the above comparison. They are /ʔ/ and /h/, which are listed for Khmu², but not for Yumbri. It seems incredible to me that Yumbri lacks /ʔ/ (glottal stop, or catch in

the throat), and unlikely that it lacks /h/, but we do not find a hint of either in Bernatzik, so far as I can interpret him. It is very likely that the word Yumbri has a glottal stop on the end [yumbriʔ] or [yimbriʔ]. Kraisi's Thai transcription implies it for Mrabri, and the cognate form of the last part of this word is /briʔ/ in Khmuʔ. It means 'wild' or 'forest'.

Using a more consistent transcription, we can now retranscribe all of Bernatzik's final consonants in phonetic or Thai characters, representing a normalized or possibly semi-phonemic analysis. The phonetic symbols are the same as I have used for transcribing Khmuʔ.

p u	t ʋ	c ɿ	k ɳ
m u	n u	ñ ɰ	ŋ ɯ
	l ʌ		
	r ʔ		
w ɿ		y ʊ	

Suggested Retranscription of Bernatzik's Final Consonants

For the initial (syllable initial) consonants we will cut short the explanation of method of analysis. We find a chart like this.

b	d	dy	g
p	t	ty	k
	ds		
	ts	tš/tš	
m	n	ny	ñ
	s	š	
w/ʉ/ǎ		y,í	

Bernatzik's Initial Consonants

ny we are assuming to be another way of writing [ñ] (Bernatzik's third way). The fact that /ʔ/ and /h/ are missing is very suspicious. That both are missing is highly improbable, but we seem to have no evidence for them in Bernatzik's data. The fact that there are no clearly indicated differences between aspirated and unaspirated stops (i.e., between /p/ ɽ and /ph/ ʋ, /t/ ʋ and /th/ ʋ, etc.) is very suspicious. It is very probable, that Bernatzik heard [th] as [ts], and [ch] as [tš]. At any rate, *ds*, *ts*, *tš*, and *š* in Bernatzik's transcription are a problem to me as I have not found them in related languages in any phonetic form which I would

transcribe with these symbols. However, they are perfectly possible and cannot be rejected *a priori*. I lean to the guess that *ds* does not differ from *ts*, and that this is Bernatzik's transcription of /th/, and that *t̄s* is his transcription of /ch/. We find no evidence for /ph/ or /kh/.

Bernatzik uses three symbols for 'w' and two for 'y', as indicated in the chart. In the word *sāw̄ūē* 'hungry' the *ū* probably represents 'w', coming as it does between two vowels. The reason for considering this a consonant rather than a vowel is technical, and we will not go into it here. In 'hunt' the *ū* of *gūā* is probably a /w/, coming as it does after a /g/. As a principle we could state that Bernatzik's *ū* or *ō* (note that they are short) coming between two long vowels or between a consonant and a long vowel should be transcribed /w/ (or ๗ in Thai script). Likewise Bernatzik's *i* in the same position should be retranscribed /y/ or ๘.

Note that in syllable initial position there is no problem about accepting both /p/ and /b/, etc. They typically contrast in this position in related languages. It was only in final position that their occurrence was suspect. Here is our retranscription.

b ๒	d ๓	j ๔	g ๕
p ๑	t ๒	c ๓	k ๔
	th ๖	ch ๗	
m ๘	n ๙	ñ ๑๐	ŋ ๑๑
	s ๑๒	š ๑๓	
	l ๑๔		
	r ๑๕		
w ๑๖		y ๑๗	

Retranscription of Bernatzik's Syllable Initial Consonants

Note that ๔ /j/, ๕ /g/, ๑๐ /ñ/, and ๑๓ /š/ are not used for their usual Thai values. They are here given the values of sounds which do not occur in Thai.

More guess work and interpolation from other languages is involved in the analysis of the vowels, and more inconsistency in Bernatzik's transcription is likely. Here is a chart of my interpretation of Bernatzik. The columns headed B contain Bernatzik's symbols. The remaining symbols are my retranscriptions in Thai and phonetic scripts.

Retranscription	B	Retranscription	B	Retranscription	B
ii - ⁴	ī	īī - ⁴ ๓		uu - ⁴	ū
i - ⁴	i	ī - ⁴	ī	u - ⁴	ǔ
	i				u
ee 1-	ē	๓๓ 1- ๓	ā	oo 1-	ō
e 1-๕	e	๓ 1- ๓๕	ǎ	o 1-๕	ǒ
					o
cc 11-	ē			๓๓ -๓	ō
c 11-๕	ē			๓ 1-๕๕	ǒ
	e				
aa -1	ā			๓๓	ō
a -๕	ǎ			๓	o
	a				

Retranscription of Bernatzik's Vowels

The Thai symbols in the chart are intended only to indicate the estimated value of the sounds for Thai readers. In retranscribing Bernatzik's material in Thai symbols the usual Thai conventions would be followed, the different symbols for the same sound being used in the appropriate places. In other words, the chart does not try to be exhaustive of the Thai *symbols* that would have to be used, but only indicates the presumed pronunciations involved in the vowel system.

Certain problems become obvious immediately. By our interpretation, for one thing, many of the short vowels are written by two symbols, one with a short mark over it, and the other without. It is possible, but unlikely, that there are really three lengths which are distinctive and

contrastive, or that there are more vowels than we are postulating, and that these additional ones do not occur long and short.

The assignment of Bernatzik's \bar{a} to /oo/ and his \bar{u} to /aa/ is pretty arbitrary. It was done on the basis of frequency. Bernatzik apparently makes a difference between \bar{a} and \bar{u} in most languages of this family have a contrast between /oo/ /aa/. /aa/ usually comes more frequently in words of this family which I have analyzed, so the assignment was made.

/ɛ̃/ and /ĩ/ (Thai $\text{เ}^{\text{า}}$ and $\text{ิ}^{\text{า}}$) are very difficult to reconstruct. When Bernatzik writes \acute{r} I suspect he means /ɛ̃r/ $\text{เ}^{\text{า}}$. Sometimes when he writes no vowel at all in a consonant cluster it should be read with /ɛ̃/. This is a common problem in languages of the family.

Bernatzik's transcription includes a set of vowels \bar{o} , \bar{u} which could be a misinterpretation on his part, or may well represent the phonetic value of an extra low back vowel, as indicated on our chart. This has no Thai counterpart. Kuy (Souey), a Mon-Khmer language of Northeast Thailand, Cambodia, South Laos, has such a vowel.

Bernatzik's $\acute{}$ accent mark probably indicates stress on the final syllable, as is common to all these languages. When it does not occur there I suspect he simply neglected to record it. Where it occurs more than once in a word, as in $\text{kr}^{\text{า}}$ 'camp' I suspect he is trying to indicate /ɛ̃/: /kɛ̃rɔɔj/ $\text{ค}^{\text{า}}$

\bar{a} is hard to interpret because it has a conflict of symbols (short and long marks) on the same letter.

With these points in mind, here are some sample retranscriptions of Bernatzik's words.

	Bernatzik	Phonetic Symbols	Thai Symbols
tongue	ɛ̃rj \bar{a} t	/atɛ̃rjaat/	อะต $\text{เ}^{\text{า}}$ ต
eye	m \bar{a} t	/maat/	มาต
dog	laš \bar{o}	/lašoo/	ละลอ
wind	rm \bar{o} t	/rɛ̃moot/	เ $\text{ิ}^{\text{า}}$ มอด