

THE RELEVANCE OF LEXICOSTATISTICS
TO MON-KHMER LANGUAGES

Franklin E. Huffman
Cornell University

In 1970-71 I collected data on fifteen Mon-Khmer languages in Thailand, Cambodia, and Laos.¹ Since the Mon-Khmer languages of Burma had been investigated by H. L. Shorto of the London School of Oriental and African Studies,² and those in Vietnam are being studied by the linguists of the Summer Institute of Linguistics,³ my research plan was to collect data on those Mon-Khmer languages in Thailand, Cambodia, and Laos for which we have inadequate or unreliable data (which would include most of them), and to work on as many of these languages as time and political conditions would permit. To this end I prepared a list of approximately 1000 items which was very carefully designed to provide information on consonants, vowels, clusters, affixation, and even syntax, being careful to include all the words of the Swadesh 200-word list⁴ for possible lexicostatistical application, as well as glosses from various published lists, such as those of Cabaton,⁵ Macey,⁶ and Crawford,⁷ for maximum comparability.

In Thailand, where travel is relatively unrestricted, I was able to collect data on the Kuy of Surin Province, the Chaobon (Niakuol)⁸ in Chayaphum Province, the Mal (Thin) in Nan Province, the Lawa (Thai /lua?/, Lawa /ləvʰə?/) of Mae Hong Sorn Province, the Chong (Chawng, Samre) in Chantaburi Province, Thai

1 from the village Bangkradi in Thonburi Province, and for comparison, Burmese Mon from a Mon refugee from Moulmein resident in Bangkok.

In Cambodia, although it was not possible to travel about the countryside, I was able to locate speakers of Tampuan, Brao, and Stieng from Ratanakiri in refugee camps on the outskirts of Phnom Penh. My great disappointment was that I was not able to contact speakers of Pear, Chong, or Moch, since I was most interested in investigating those languages which are presumed to be closest to Cambodian itself, *i.e.* in Thomas and Headley's terms, the Pearic group.⁹

In Laos, it turned out that the misfortunes of my work were in a sense my good fortune, since in Pakse I was able to find among the students of the Collège Saravane and the Collège d'Attapeu, which had been evacuated to Pakse, speakers of a half-dozen Mon-Khmer languages. I decided to work on Loven, Souei, Harlak (/harlak/), and Nge? (Ngeh, Kriang) as representative of the four linguistically most diverse groups of dialects. Finally, in Savannakhet I reported a speaker of Bru (So)¹⁰ from a refugee camp east of Savannakhet into town for a week, since it was not considered safe to sleep in the camp.

The time required to collect the body of data and make a rudimentary phonological analysis for each language, following the format of eliciting data from the informant in the morning and collating and analyzing in the afternoon, decreased from about three weeks per language in the beginning to about a week for each of the languages in Laos, partly, I suspect, because working conditions were less congenial than had been my hotel room in Bangkok where I

had worked on Burmese Mon, Thai Mon, and Kuy, but more importantly because experience with a number of *structurally similar languages enabled me progressively* to refine and formalize techniques of both elicitation and analysis. It was nevertheless very reassuring to later discover that my analysis of Lawa was virtually identical with that of Schlatter,¹¹ differing from his only in lacking one of his diphthongs, which was apparently introduced by Thai loans, and in finding an initial preaspirated nasal in the palatal position, which Schlatter's analysis does not show, to fill out the preaspirated nasal series.

Thus I had managed, in a period of about nine months, to collect a body of data and make a rough phonological analysis of fifteen languages. The question was, what did I have, and what could I do with it? Now, it became apparent quite early on, perhaps as early as the Cambodian phase of the project, that my research design was faulty and that my results were going to be neither fish nor fowl--in short, that I was going into these languages in greater detail than was really necessary for a simple lexicostatistical survey à la Swadesh, but on the other hand in insufficient depth for the purposes of traditional comparative reconstruction. Balanced against this realization, however, was the desire to get as much data as I could on as many languages as I could in the time available, since it was not at all clear when anyone would have access to some of these languages again, and especially since these languages are absorbing vocabulary, not to mention grammatical elements, from surrounding regional or national languages, whether Thai, Lao, or Cambodian, at a rapid rate, and some of them are on the very

think of being replaced entirely. For example, a rough count shows that over a list of 1,000 words of basic vocabulary the Mon-Khmer languages in Thailand and Laos have borrowed about 200 words, or 20 percent, from Thai or Lao; the percentage of loans in the total lexicon would presumably be much higher. In fact, displacement is so pervasive that I was not able to come up with even a 100-word list for which native Mon-Khmer words had been retained straight across all fifteen languages. For some reason the languages investigated in Cambodia show a smaller percentage of borrowed vocabulary, less than five percent over 1,000 words, perhaps because Khmer loans are harder to identify than are Lao, Thai, or Vietnamese loans.

The inadequacy of my research design was due in part to certain false assumptions on my part. Since one of the purposes of research is to reveal how and why our assumptions or hypotheses are wrong, it might be instructive to look at some of them.

One such assumption concerned the linguistic distance between the languages to be compared. It occurred to me that it might be a good idea to organize at least part of my questionnaire in groups of words having the same vowel in Cambodian, resulting in 31 groups of words based on the 31 vocalic nuclei of modern standard Cambodian.¹² My hope was that for a given group of ten words having the same vowel in Cambodian, another language X would have a constant vowel, whether i, e, u, o, or whatever, in a significant number of the same words, thus revealing certain vowel-shift patterns as well as some of the conditioning factors involved. But given the high degree of lexical displacement, or simple linguistic

distance, between Mon-Khmer languages, this plan was doomed to failure. For example, for the words 'clean', 'far', 'fear', 'follow', 'get', 'head', 'light (in weight)', 'say', 'tongue', and 'mean', all having long /aa/ in Cambodian, Loven (to cite only one) has /luh/, /hɣaay/, /hmo?/, /tooy/, /bic/, /tuus/, /hjaac/, /moh/, /piat/, and /suə/, a different vowel in each word and all except /hɣaay/ 'far' unrelated to the Cambodian forms. Thus it would have made the task of elicitation much simpler, given the diversity of the languages, to have organized the corpus in groups of semantically related items, such as parts of the body, numbers, kinship terms, and antonyms such as 'hot:cold', 'large:small'.

Another experiment which failed was the assumption that a significant amount of morphological and syntactic information could be elicited from a pre-established corpus constructed by analogy with Cambodian structure. Although I had no great hope for this assumption from the outset, it became immediately clear that each language must be investigated in terms of its own system, and that to get a complete picture of that system would have required a much more intensive investigation of each language.

The two terms glottochronology and lexicostatistics have come to be used almost interchangeably, although their assumptions are, I believe, quite different. Glottochronology, as formulated by Swadesh,¹³ made the following assumptions:

1. that items of basic vocabulary are replaced at a constant rate in all languages, and
2. that a universally applicable list of basic vocabulary can be devised for all languages.

th of these assumptions have by now been discredited.¹⁴ Much more acceptable to most of us is the theory that language change can be radically affected by density of contact, and by relative status of the recipient and donor languages. This is readily apparent in the effects of increased density of mass communication media, as well as teaching of standard national languages in the schools, on minority and tribal Mon-Khmer languages in Southeast Asia.

As for the second assumption, Swadesh's original list of 200 words¹⁵ was immediately attacked by specialists in various language families as being inapplicable. To test the truth of this charge we need only consider several of the items on this list in relation to Southeast Asian languages in general. To begin with, 'brother' won't work, since there is no general term for brother in any Southeast Asian language I know of, siblings always being specified by relative age. Even worse, consider such general concepts as 'to cut', 'to hit', or 'to wash'. Given the verbal specialization characteristic of Asian languages, any researcher who goes to Southeast Asia expecting to elicit a general term for 'to cut' is doomed to failure. Southeast Asians never cut; they slice, chop, hack, snip, mince, or split, but they seldom, if ever, cut. Thus beleaguered, adherents of glottochronology culled their list down to 100 items.¹⁶ But even the 100-word list has its problems; e.g. it retains the pronouns 'I', 'thou', and 'we', which cause problems in any Southeast Asian language. Thus the comment of G. L. Trager regarding glottochronology seems valid: 'The notion of an universal vocabulary, no matter how small, is

nonsense. Languages are not merely lists of words; they are systems.'¹⁷

Hymes, however, in 1960¹⁸ made a terminological distinction between glottochronology and lexico-statistics which appears to me to be useful. He defined lexicostatistics as the simple quantification of cognates sharing a common gloss, without the historical implications of glottochronology. Thus word lists could be tailored to the cultural and semantic specifications of a given group of languages, thereby avoiding the assumption of universality of basic vocabulary. I would agree with Thomas and Headley's statement: "Lexicostatistics is not a precision tool. Careful phonological reconstruction is necessary if one desires detailed information about language relationships. Lexicostatistics is useful, however, for giving a quick general picture of language groupings."¹⁹

Given this definition of lexicostatistics, then, it occurred to me that a much larger list of words than the Swadesh 200, freed from the requirement of universality except with regard to the specific group of languages being compared, should provide much more accurate and definitive information on sub-grouping than would a smaller sample, since presumably more closely related languages would share a much higher percentage of relatively unstable cultural vocabulary, thus providing a more highly calibrated spectrum of differentiation. So I set out to select the largest possible list of acceptable items from my corpus to serve as the largest common denominator for comparison. I soon found that large numbers of individual items had to be rejected for one of the following reasons:

1. Item missing in several of the 15 languages. Items of this kind were either not culturally relevant, such as 'bear', 'tiger', 'yoke', 'shield', or were too abstract, such as 'game', 'war', or 'peace'.

2. Item has same translation as another item. For example, it is pointless in most Mon-Khmer languages to include both 'foot' and 'leg', 'arm' and 'hand', or 'day' and 'sun'.

3. Item typically translated by a compound using words already included. Examples of this are 'tree' (talk-wood), 'beard' (hair-chin), and 'milk' (water-east).

4. Item too general to elicit a consistent equivalent across languages, such as 'bad', 'good', 'basket', or 'worm'.

5. Item is typically a loan word, such as 'book', 'bottle', or 'market'. With loan words, however, it was frequently difficult to decide whether an item had been borrowed in common from Thai languages by all the Mon-Khmer languages involved, or whether the word had originally been borrowed into Thai from Cambodian and later passed on to various Mon-Khmer languages through Thai and Lao. Consider, for example, the word for 'candle', which is /tʰien/ in Cambodian and /thian/ in Thai and Lao. The characteristic uniformity of the form across almost all the languages suggests that it is in fact a recent loan. On the other hand, the form of the word Loven, which retains the original voiced series /d j g/, is /dian/, which is precisely what one would expect. If the word had been borrowed into Loven from Lao, one would expect the form /hian/, since Loven also has a series of

aspirated initial consonants. Further evidence is the Vietnamese form /dèn/ 'candle'. Another problematic item is the word for 'skillet', which is /kteǎh/ in Cambodian, /kratha?/ in Thai, and /katha?/ in Lao. Given the characteristically Cambodian initial cluster and final aspiration, this word is obviously borrowed into Thai and Lao from Cambodian. The Chaobon and Kuy forms /kəta?/ and the Bru form /kətha?/ are just as obviously borrowed in turn from Thai and Lao. On the other hand, Alak /k'at'ah/, Tampuon /kət'ah/, and Chong /tǎh/ retain final aspiration, which suggests that they are linear rather than second-hand cognates. Other examples of problematic loan words are 'believe' (Camb. /cuəy/, Thai /chuâj/) and 'be born' (Camb. /kaet/, Thai /kəet/). It was decided to retain such problematic items for the list pending further evidence. It was also decided to retain certain items for which several languages had obvious loans, since replacement from whatever source is presumably grist for the lexicostatistical mill, so long as the item provided valuable information for some languages, and with the stipulation that common borrowing would be treated as non-cognate.

Using these criteria, I found I had pared my corpus of acceptable items by about 50 percent, so I decided, for ease of calculation, to hold the line at 500 words, for better or for worse. To the 15 languages I already had, I added data from Cambodian, Vietnamese, Pear from Morizon,²⁰ and Khmu? from Smalley.²¹ Of my 500-word list I was able to find only 353 items for Pear and 207 for Khmu?, so that in figuring percentages involving these two languages the denominators had to be adjusted accordingly.

The number of possible pairs in a set of 19 languages is $n \times (\frac{n-1}{2})$, or 171. The task of comparing 500 items 171 times turned out to be gargantuan, and the results disappointing. The highest percentage of cognation was 86 between the Burmese and Thai dialects of Mon, which are dialects of the same language. In fact, my informants for the two dialects were able to communicate with each other after some initial adjustment. The next highest percentages fell to around 50, between Kuy of Thailand and Souei and Bru of Laos, all considered by Thomas and Headley to belong to the Katuic group, and between Pear and Chong of Thomas and Headley's family.²² Sixteen more percentages range between 40 and 31, while all the remaining 150 percentages range between 27 and 5, with Khmu?, Mal, Chaobon, Hmong, Lawa, and Vietnamese showing the lowest average percentages. In other words, if we treat the Burmese and Thai dialects of Mon as a single language, all the percentages fall between 55 and 5, with the great majority clustering in the teens and twenties. My somewhat chagrined conclusion was that, for languages so distantly related as these, a 500-word list was rather all too large, and contained too many relatively unstable items, to reveal significant sub-groupings.

Unhappy with such modest results from so much labor, I decided out of curiosity to compare the cognate percentages based on a carefully chosen list of 100 items of hard-core vocabulary. I had ample time, in the course of my 171 peregrinations through the 500-word list, to observe that there was a small core of items that was highly persistent, with perhaps one or two exceptions each, across all 19

languages. Since there are only about 30 of them, it might be of some interest to list them, with a kind of canonical form, or phonological average, for each:

- | | |
|--------------------------|------------------|
| 1. one /muəy/ | 16. earth /tɛʔ/ |
| 2. two /baar/ | 17. eye /mat/ |
| 3. three /pee/ | 18. fish /kaa/ |
| 4. four /poon/ | 19. a fly /ruy/ |
| 5. child /koon/ | 20. foot /juŋ/ |
| 6. grandchild /caw/ | 21. hair /sok/ |
| 7. ant /mooc/ | 22. hand /tii/ |
| 8. bird /ceem/ | 23. horse /seh/ |
| 9. bitter /taŋ/ | 24. ivory /blok/ |
| 10. breast /tɔh/ | 25. louse /cay/ |
| 11. to cry /yaam/ | 26. nose /muh/ |
| 12. day /ŋay/ | 27. root /rɛh/ |
| 13. dog /cɔɔ/ | 28. shoot /pañ/ |
| 14. done (cooked) /ciin/ | 29. water /daak/ |
| 15. duck /daa/ | 30. weave /taañ/ |

Of these 30 items, 10 are not included in the Swadesh 200-word list, which suggests that stable vocabulary is highly specific to individual language families. I decided to include these 30 words on my 100-word list, to provide a fairly predictable base percentage for the highly marginal languages such as Mon, Lawa, and Vietnamese.

At the same time, I noticed that certain presumably highly stable items showed unexpected diversity. The following 20 items, all of which are included on the Swadesh 200-word list, show great diversity: black, cold, come, ear, fire, give, head, husband, I, large, mountain, mouth, say, short, skin, small, tail, tooth, wife, and walk. A rather puzzling example of this diversity is the item

head', which is Cambodian /kbaal/, Pear *tos* (sic
 om Morizon), Chong /təot/, Tampuon /kal/, Stieng
 ook/, Brao and Loven /tuux/, Alak /k'ɬɬ/, Bru,
 e?, and Souei /pləə/, Kuy /plɔɔ/, Mal /k'ɬ?/,
 aobon /kədap/ Mon /dɔp/, Lawa /kañ/, and Viet-
 mese *dầu* (/dèw/). My colleague Philip Jenner has
 suggested that the diversity here might be due to
 placement resulting from the head taboo common in
 Southeast Asian cultures.

Having added these 20 items to the list to high-
 light highly specialized resemblances, I made up the
 remaining 50 items from words whose cognates fell
 to significant groups, thus providing pivotal
 information on cleavages. Examples are 'five' with
 ram/ or /sɔɔŋ/ forms, 'water buffalo' with
 rapaw/, /traak/, or /priaŋ/ forms, 'chicken' with
 truey/ or /ʔier/ forms, 'snake' with /kəsañ/ or
 ih/ forms, and 'tongue' with /ntaak/ or /mpiət/
 forms. Only 94 of the 100-word list were included in
 Morizon's Pear data, and only 63 could be found in
 Alley's Khmu? data. Although the denominators were
 adjusted accordingly in figuring percentages for
 these languages, the Khmu? list should probably be
 considered too small to be reliable.

Having thus very scientifically composed my 100-
 word list, I proceeded to again calculate cognate
 percentages 171 times, based on this 100-word list.
 In treating Burmese Mon and Thai Mon as a single
 language, since they showed 99 percent cognation with
 this list, all the percentages fell within the range
 80 to 20, and while being a bit more diffuse in
 distribution than the 500-word percentages, tended to
 cluster in the 45 to 25 percent range.

Thus while the distribution pattern remained roughly comparable, the percentages based on the 100-word list were an average of 22.7 percentage points higher than those based on the 500-word list. This finding suggests that for any given set of languages, the cognate percentages will be inversely proportional to the size of the list used (assuming the most basic vocabulary for each list). This conclusion is strengthened by the fact that the percentages obtained by Thomas and Headley,²³ who used a 207-word list--*i.e.* roughly intermediate between my 500-word and 100-word lists--are roughly intermediate between my figures for the same pairs of languages.

Given this fact, Swadesh's dictum that percentages of 100-81 indicate dialects of one language, 81-36 members of one family, and 36-12 members of a common stock²⁴ is obviously unreliable.

Percentages can further be manipulated by prejudicing one's list in favor of a particular language. For example, from my 500-word list I was able to identify approximately 75 Vietnamese words as apparent Mon-Khmer cognates. Only 40 of these happened to be included in the 100-word list, but if I had arbitrarily included all 75 of these items in the list, the Vietnamese percentages would have been unnaturally high in relation to those of other languages; and if I had further narrowed the test list to these 75 words, Vietnamese might have shown almost 100 percent cognation with some languages, rather than the 33 to 21 percent range actually obtained with the 100-word list, or the 11 to 5 percent range obtained with the 500-word list.

As a parting shot at lexicostatistics, I decided average the 500-word and 100-word percentages obtained for each of the 171 pairs of languages. Again, or rather, inevitably, the distribution pattern of percentages was comparable to that of the 50- and 100-word percentages. These averages are furthermore very close to the percentages obtained by Thomas and Headley, and in general tend to support their conclusions, although we may be able to reclassify certain languages in the light of fuller evidence. According to my figures, for example, Alak, which Thomas and Headley include in the Katuic branch, is closer to Loven and Brao of West Bahnaric than to Nge?, Bru, Souei, or Kuy of Katuic; and is closer to Stieng of South Bahnaric than to Bru, Souei, or Kuy of Katuic. It appears, then, that Alak could be reclassified as West Bahnaric.²⁵

My conclusions, which seem rather trite, are that 1. basic vocabulary is highly specific to individual groups of languages, or perhaps to individual cultures, and that 2. lexicostatistics, as defined by Hymes, is useful in showing relative distance between languages within a given group of languages and using a given corpus of vocabulary, but that absolute percentages are meaningless.

A chart showing the 500-word and 100-word percentages and their averages for each of the 171 pairs of languages, arranged in descending order of relationship, is appended.

APPENDIX

COGNATE PERCENTAGES OF 171 PAIRS OF MON-KHMER LANGUAGES

500-Word List		100-word List		Average	
Cambodian					
Pear	37	Stien	50	Stien	42
Stien	34	Kuy	49	Pear	40
Kuy	31	Brao	47	Kuy	40
Brao	31	Tamp	47	Brao	39
Tamp	27	Pear	44	Tamp	37
Chong	26	Alak	44	Alak	33
Chaob	26	Loven	43	Loven	32
Souei	23	Chaob	41	Chong	32
Alak	23	Nge?	40	Chaob	32
Bru	21	Chong	39	Souei	31
Loven	21	Souei	39	Bru	30
Nge?	19	Bru	39	Nge?	29
Khmu	16	T Mon	33	T Mon	23
Lawa	14	B Mon	33	B Mon	22
T Mon	13	Lawa	31	Lawa	22
B Mon	13	Viet	29	Khmu	21
Mal	12	Khmu	27	Viet	20
Viet	11	Mal	25	Mal	18

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

00-word List		100-word List		Average	
Pear					
Chong	54	Chong	76	Chong	65
Camb	37	Camb	44	Camb	40
Kuy	22	Loven	44	Loven	32
Stien	22	Brao	43	Brao	31
Loven	20	Alak	40	Stien	30
Brao	19	Stien	38	Kuy	30
Tamp	18	Kuy	38	Alak	29
Alak	18	Bru	38	Tamp	27
Souei	18	Tamp	37	Bru	27
Bru	16	Nge?	36	Souei	26
Nge?	15	Souei	35	Nge?	25
Chaob	15	Lawa	31	Chaob	22
Lawa	12	Chaob	30	T Mon	18
Khmu	11	Viet	28	B Mon	18
Mal	10	T Mon	28	Khmu	18
B Mon	10	B Mon	27	Viet	18
T Mon	9	Khmu	25	Mal	17
Viet	8	Mal	24	Lawa	16

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word list		Average	
Chong					
Pear	54	Pear	76	Pear	65
Camb	26	Camb	39	Camb	32
Stien	18	Stien	36	Stien	27
Tamp	18	Brao	36	Brao	27
Brao	18	Loven	35	Tamp	26
Kuy	18	Tamp	34	Kuy	25
Alak	16	Kuy	33	Loven	25
Souei	16	Alak	33	Alak	24
Loven	15	Souei	31	Souei	23
Nge?	14	Bru	30	Bru	22
Bru	14	Nge?	29	Nge?	21
Chaob	14	Chaob	27	Chaob	20
Khmu	12	Lawa	25	Lawa	17
Lawa	9	T Mon	23	Khmu	16
Mal	9	B Mon	22	Mal	15
T Mon	8	Mal	22	T Mon	15
B Mon	7	Viet	22	B Mon	14
Viet	6	Khmu	21	Viet	14

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Stieng					
Camb	34	Tamp	63	Tamp	48
Tamp	33	Brao	58	Brao	45
Brao	32	Alak	56	Camb	42
Loven	25	Loven	54	Alak	40
Alak	25	Kuy	52	Loven	39
Kuy	25	Camb	50	Kuy	38
Souei	23	Souei	46	Souei	34
Pear	22	Bru	46	Bru	33
Nge?	20	Chaob	44	Nge?	31
Bru	20	Nge?	42	Pear	30
Chong	18	Pear	38	Chaob	29
Chaob	15	T Mon	38	Chong	27
Khmu	12	B Mon	38	B Mon	24
Lawa	10	Chong	36	T Mon	24
Mal	10	Lawa	32	Lawa	21
T Mon	10	Khmu	29	Khmu	20
B Mon	10	Viet	29	Mal	18
Viet	7	Mal	26	Viet	18

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Tampuon					
Brao	34	Brao	64	Brao	49
Stien	33	Stien	63	Stien	48
Camb	27	Alak	63	Alak	45
Alak	27	Loven	62	Loven	44
Loven	26	Kuy	53	Kuy	37
Kuy	22	Nge?	53	Camb	37
Bru	21	Bru	52	Bru	36
Souei	20	Souei	52	Souei	36
Nge?	19	Camb	47	Nge?	36
Pear	18	Chaob	41	Pear	27
Chong	18	T Mon	41	Chong	26
Khmu	12	B Mon	40	Chaob	26
Chaob	12	Pear	37	B Mon	25
T Mon	10	Lawa	36	T Mon	25
B Mon	10	Chong	34	Lawa	22
Lawa	9	Viet	30	Khmu	19
Mal	8	Khmu	27	Viet	18
Viet	7	Mal	26	Mal	17

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List	100-word List	Average
Brao		
Loven 44	Loven 77	Loven 60
Tamp 34	Tamp 64	Tamp 49
Alak 33	Alak 63	Alak 48
Stien 32	Kuy 61	Stien 45
Camb 31	Stien 58	Kuy 43
Kuy 25	Souei 54	Camb 39
Bru 24	Bru 52	Souei 38
Souei 23	Nge? 51	Bru 38
Nge? 22	Camb 47	Nge? 36
Pear 19	Chaob 44	Pear 31
Chong 18	Pear 43	Chaob 30
Chaob 16	T Mon 42	Chong 27
Khmu 14	B Mon 41	T Mon 26
Lawa 11	Chong 36	B Mon 25
T Mon 10	Lawa 36	Khmu 23
B Mon 10	Khmu 32	Lawa 23
Mal 10	Mal 29	Mal 19
Viet 7	Viet 28	Viet 17

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Loven					
Brao	44	Brao	77	Brao	60
Alak	38	Alak	69	Alak	58
Nge?	27	Tamp	62	Tamp	44
Souei	27	Souei	60	Souei	43
Tamp	26	Bru	57	Bru	41
Stien	25	Kuy	55	Kuy	39
Bru	25	Stien	54	Nge?	39
Kuy	24	Nge?	52	Stien	39
Camb	21	T Mon	47	Camb	32
Pear	20	B Mon	46	Pear	32
Chong	15	Pear	44	Chaob	28
Chaob	14	Camb	43	T Mon	28
Khmu	14	Chaob	43	B Mon	28
Mal	11	Lawa	36	Chong	25
Lawa	11	Chong	35	Khmu	23
B Mon	10	Mal	34	Lawa	23
T Mon	10	Khmu	32	Mal	22
Viet	7	Viet	29	Viet	18

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Alak					
Loven	38	Loven	69	Loven	58
Brao	33	Brao	63	Brao	48
Nge?	31	Tamp	63	Tamp	45
Tamp	27	Stien	56	Nge?	42
Souei	26	Nge?	53	Stien	40
Stien	25	Bru	52	Bru	38
Bru	25	Kuy	51	Souei	37
Kuy	23	Souei	49	Kuy	37
Camb	23	T Mon	47	Camb	33
Pear	18	B Mon	46	Pear	29
Chong	16	Chaob	45	Chaob	29
Khmu	14	Camb	44	T Mon	27
Chaob	13	Pear	40	B Mon	27
Mal	12	Lawa	36	Lawa	24
Lawa	12	Mal	34	Chong	24
T Mon	8	Chong	33	Mal	23
B Mon	8	Khmu	30	Khmu	22
Viet	6	Viet	30	Viet	18

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Nge?					
Bru	43	Bru	66	Bru	54
Souei	40	Souei	65	Souei	52
Kuy	35	Kuy	64	Kuy	49
Alak	31	Alak	53	Alak	42
Loven	27	Tamp	53	Loven	39
Brao	22	Loven	52	Tamp	36
Stien	20	Brao	51	Brao	36
Tamp	19	Stien	42	Stien	31
Camb	19	T Mon	41	Camb	29
Pear	15	B Mon	40	Pear	25
Chong	14	Camb	40	Chaob	25
Chaob	13	Chaob	38	T Mon	25
Khmu	13	Pear	36	B Mon	24
Mal	12	Lawa	32	Khmu	24
Lawa	11	Chong	29	Chong	21
T Mon	9	Mal	29	Lawa	21
B Mon	8	Viet	28	Mal	20
Viet	7	Khmu	25	Viet	17

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Bru					
Souei	55	Souei	82	Souei	68
Kuy	49	Kuy	75	Kuy	62
Nge?	43	Nge?	66	Nge?	54
Loven	25	Loven	57	Loven	41
Alak	25	Alak	52	Brao	38
Brao	24	Brao	52	Alak	38
Tamp	21	Tamp	52	Tamp	36
Camb	21	Stien	46	Stien	33
Stien	20	Chaob	43	Camb	30
Chaob	16	T Mon	42	Chaob	29
Pear	16	B Mon	41	Pear	27
Khmu	15	Camb	39	T Mon	26
Chong	14	Pear	38	B Mon	25
Mal	13	Viet	33	Chong	22
Lawa	12	Lawa	32	Khmu	22
T Mon	11	Khmu	30	Lawa	22
B Mon	10	Chong	30	Mal	21
Viet	8	Mal	30	Viet	20

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Souei					
Bru	55	Bru	82	Bru	68
Kuy	54	Kuy	81	Kuy	67
Nge?	40	Nge?	65	Nge?	52
Loven	27	Loven	60	Loven	43
Alak	26	Brao	54	Brao	38
Brao	23	Tamp	52	Alak	37
Stien	23	Alak	49	Tamp	36
Camb	23	Stien	46	Stien	34
Tamp	20	Chaob	44	Camb	31
Chaob	18	T Mon	40	Chaob	31
Pear	18	B Mon	39	Pear	26
Chong	16	Camb	39	T Mon	26
Lawa	13	Pear	35	B Mon	25
Mal	13	Lawa	34	Chong	23
Khmu	13	Mal	32	Lawa	23
T Mon	12	Viet	32	Mal	22
B Mon	11	Chong	31	Khmu	21
Viet	7	Khmu	30	Viet	19

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Kuy					
Souei	54	Souei	81	Souei	67
Bru	49	Bru	75	Bru	62
Nge?	35	Nge?	64	Nge?	49
Camb	31	Brao	61	Brao	43
Brao	25	Loven	55	Camb	40
Stien	25	Tamp	53	Loven	39
Loven	24	Stien	52	Stien	38
Alak	23	Alak	51	Tamp	37
Tamp	22	Camb	49	Alak	37
Pear	22	Chaob	44	Chaob	32
Chaob	20	T Mon	43	Pear	30
Chong	18	B Mon	42	T Mon	28
Khmu	16	Pear	38	B Mon	27
Lawa	14	Mal	36	Chong	25
T Mon	14	Lawa	34	Khmu	24
B Mon	13	Chong	33	Mal	24
Mal	12	Khmu	32	Lawa	24
Viet	8	Viet	32	Viet	20

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Khmu					
Mal	18	Mal	41	Mal	29
Lawa	18	Lawa	33	Lawa	25
Kuy	16	Kuy	32	Kuy	24
Camb	16	Brao	32	Nge?	24
Bru	15	Loven	32	Brao	23
Brao	14	Bru	30	Loven	23
Loven	14	Alak	30	Bru	22
Alak	14	Souei	30	Alak	22
Souei	13	Stien	29	Camb	21
Nge?	13	Camb	27	Souei	21
Stien	12	Tamp	27	Stien	20
Tamp	12	Nge?	25	Tamp	19
Chong	12	Pear	25	Pear	18
Pear	11	Chaob	22	Chong	16
Chaob	11	Viet	22	Chaob	16
T Mon	8	Chong	21	T Mon	13
B Mon	7	T Mon	19	B Mon	13
Viet	5	B Mon	19	Viet	13

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Mal					
Khmu	18	Khmu	41	Khmu	29
Souei	13	Kuy	36	Kuy	24
Bru	13	Alak	34	Alak	23
Kuy	12	Loven	34	Loven	22
Alak	12	Souei	32	Souei	22
Nge?	12	Bru	30	Bru	21
Chaob	12	Nge?	29	Chaob	20
Camb	12	Brao	29	Nge?	20
Loven	11	Chaob	28	Brao	19
Lawa	11	T Mon	28	Stien	18
Brao	10	B Mon	28	Camb	18
Stien	10	Lawa	26	Lawa	18
Pear	10	Stien	26	T Mon	18
Chong	9	Tamp	26	B Mon	18
T Mon	8	Camb	25	Pear	17
B Mon	8	Pear	24	Tamp	17
Tamp	8	Chong	22	Chong	15
Viet	8	Viet	21	Viet	14

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List	100-word List	Average
---------------	---------------	---------

Chaobon

T Mon 33	T Mon 69	T Mon 51
B Mon 31	B Mon 69	B Mon 50
Camb 24	Alak 45	Camb 32
Kuy 20	Kuy 44	Kuy 32
Souei 18	Souei 44	Souei 31
Brao 16	Brao 44	Brao 30
Bru 16	Stien 44	Bru 29
Stien 15	Bru 43	Alak 29
Pear 15	Loven 43	Stien 29
Loven 14	Camb 41	Loven 28
Chong 14	Tamp 41	Tamp 26
Alak 13	Nge? 38	Nge? 25
Nge? 13	Viet 31	Pear 22
Lawa 13	Pear 30	Lawa 21
Tamp 12	Lawa 29	Chong 20
Mal 12	Mal 28	Mal 20
Khmu 11	Chong 27	Viet 18
Viet 6	Khmu 22	Khmu 16

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Burmese Mon					
T Mon	86	T Mon	99	T Mon	92
Chaob	31	Chaob	69	Chaob	50
Kuy	13	Loven	46	Loven	28
Camb	12	Alak	46	Kuy	27
Souei	11	Kuy	42	Alak	27
Loven	10	Brao	41	Bru	25
Brao	10	Bru	41	Souei	25
Bru	10	Tamp	40	Brao	25
Tamp	10	Nge?	40	Tamp	25
Stien	10	Souei	39	Nge?	24
Lawa	10	Stien	38	Stien	24
Pear	10	Camb	33	Camb	22
Alak	8	Lawa	33	Lawa	21
Nge?	8	Mal	28	Pear	18
Mal	8	Pear	27	Mal	18
Chong	7	Viet	27	Viet	16
Khmu	7	Chong	22	Chong	14
Viet	6	Khmu	19	Khmu	13

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Thai Mon					
B Mon	86	B Mon	99	B Mon	92
Chaob	33	Chaob	69	Chaob	51
Kuy	14	Loven	47	Loven	28
Camb	13	Alak	47	Kuy	28
Souei	12	Kuy	43	Alak	27
Bru	11	Bru	42	Bru	26
Lawa	11	Brao	42	Souei	26
Loven	10	Tamp	41	Brao	26
Brao	10	Nge?	41	Tamp	25
Tamp	10	Souei	40	Nge?	25
Stien	10	Stien	38	Stien	24
Nge?	9	Lawa	34	Camb	23
Pear	9	Camb	33	Lawa	22
Alak	8	Pear	28	Pear	18
Mal	8	Mal	28	Mal	18
Chong	8	Viet	28	Viet	17
Khmu	8	Chong	23	Chong	15
Viet	8	Khmu	19	Khmu	13

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Lawa					
Khmu	18	Alak	36	Khmu	25
Kuy	14	Loven	36	Kuy	24
Camb	14	Brao	36	Alak	24
Souei	13	Tamp	36	Souei	23
Chaob	13	Kuy	34	Loven	23
Alak	12	Souei	34	Brao	23
Bru	12	T Mon	34	Camb	22
Pear	12	B Mon	33	Tamp	22
Brao	11	Khmu	33	Bru	22
Loven	11	Bru	32	T Mon	22
Nge?	11	Nge?	32	B Mon	21
Mal	11	Stien	32	Chaob	21
T Mon	11	Camb	31	Nge?	21
B Mon	10	Pear	31	Stien	21
Stien	10	Chaob	29	Mal	18
Tamp	9	Mal	26	Chong	17
Chong	9	Chong	25	Pear	16
Viet	5	Viet	23	Viet	14

COGNATE PERCENTAGES OF 171
PAIRS OF MON-KHMER LANGUAGES

500-word List		100-word List		Average	
Vietnamese					
Camb	11	Bru	33	Camb	20
Bru	8	Kuy	32	Kuy	20
Kuy	8	Souei	32	Bru	20
Pear	8	Chaob	30	Souei	19
Mal	8	Tamp	30	Pear	18
Souei	7	Alak	30	Stien	18
Tamp	7	Camb	29	Tamp	18
Stie	7	Stie	29	Loven	18
Loven	7	Loven	29	Alak	18
Brao	7	Pear	28	Chaob	18
Nge?	7	Brao	28	Brao	17
Chaob	6	Nge?	28	Nge?	17
Alak	6	T Mon	28	T Mon	17
T Mon	6	B Mon	27	B Mon	16
B Mon	6	Lawa	23	Chong	14
Chong	6	Chong	22	Mal	14
Lawa	5	Khmu	22	Lawa	14
Khmu	5	Mal	21	Khmu	13

¹The research on which this paper is based was supported by a fellowship from the John Simon Guggenheim Memorial Foundation, and by a travel and research grant from the National Science Foundation.

²See, for example, H. L. Shorto, *A Dictionary of Modern Spoken Mon* (London: Oxford University Press, 1962); "Word and Syllable Patterns in Palaung," in *SOAS*, 23 (1960): 544-57; "The Structural Patterns of Northern Mon-Khmer Languages," in *Linguistic Comparison in South-East Asia and the Pacific* (London: SOAS, 1963); "Mon Vowel Systems: A Problem in Phonological Statement", in C. E. Bazell, ed., *Memory of J. R. Firth* (London: Longmans, 1966), 38-409.

³See, for example, Banker et al., *Mon-Khmer Studies I* (Saigon: Summer Institute of Linguistics, 1964); Cohen et al., *Mon-Khmer Studies II* (Saigon: Summer Institute of Linguistics, 1965); Johnston et al., *Mon-Khmer Studies III* (Saigon: Summer Institute of Linguistics, 1969). A fourth volume, *Mon-Khmer Studies IV*, is anticipated from the Center for Vietnamese Studies at Carbondale, Illinois.

⁴Morris Swadesh, "Lexico-statistic dating of prehistoric ethnic contacts", in *Proceedings of the American Philosophical Society*, 96 (1952): 452-63.

⁵Antoine Cabaton, "Dix dialectes indochinois cueillis par Prosper Odend'hal. Étude linguistique de Antoine Cabaton," in *JA*, 10^e sér., 5 (1905): 335-344.

⁶Paul Macey, "Étude ethnographique sur diverses tribus aborigènes et autochtones habitant les provinces des Hua-phan[s] Ha-tang-hoc et du Cammon, Laos", in *Int. Congr. of Orient.*, 14.1 (1906): 3-63.

⁷John Crawford, *Journal of an Embassy from the Governor-General of India to the Courts of Siam and Szechin China* (London: Henry Colburn, 1828) [reprinted with an Introduction by David K. Wyatt, New York: Oxford University Press, 1967]].

⁸I studied the dialect of Ban Wang Ai Pho, Mueang Narong District, Chayaphum Province, a mountain village on the western edge of the Northeast plateau. The people call themselves /ñah kur/, 'mountain people'; although they speak a slightly different dialect, these people are obviously closely related to the Niakuol referred to by Seidenfaden in several articles in *JSS* about 50 years ago [*viz.*, Major E. Seidenfaden, "Some Notes about the Chaubun:

A Disappearing Tribe in the Korat Province", *JSS*, 12 (1918).3: 1-11; "Further Notes About the Chaubun, Etc.", *JSS*, 13 (1919).3: 47-53]. These people have also been erroneously referred to as Lawa; cf. Phra Petchabūnuri, "The Lawā or Chaubun in Changvad Petchabūn", *JSS*, 14 (1921).1: 19-45.

⁹David Thomas and Robert K. Headley, Jr., "More on Mon-Khmer Subgroupings", in *Lingua*, 25 (1970): 398-418.

¹⁰My Bru vocabulary is virtually identical with that collected by Ferlus in Sayaburi, Laos, and listed in Haudricourt's "Mutation consonantique en Mon-Khmer" [*BSLP*, 60 (1965): 160-72]. Ferlus furthermore states [in "Ou en est l'Atlas ethnolinguistique?", in *Bulletin du Centre de Documentation et de Recherche: Asie du Sud-Est et Monde Insulindien* (CeDRASEMI), fasc. 1, Vol. II, no. 4 (Dec. 1971): 74] that "les Sô du Laos se disent Bru".

¹¹Donald Schlatter, "Lawa", in William A. Smalley, ed., *Phonemes and Orthography of Eight Marginal Languages of Thailand*, Chapter 6. [Not yet published; I am indebted to Bill Smalley for prepublication drafts of this chapter, as well as David Filbeck's chapter on "Mal (Thin)" and Beulah M. Johnston's chapter on "Kuy (Suay)"].

¹²See Franklin E. Huffman, *Cambodian System of Writing and Beginning Reader* (New Haven: Yale University Press, 1960), Chapter II, "Phonology", 6-12; *Outline of Cambodian Grammar*, Ph. D. dissertation, Cornell University, 1967 (available from University Microfilms, Inc., Ann Arbor, Michigan), 244-6.

¹³Morris Swadesh, "Diffusional cumulation and archaic residue as historical explanation", in *Southwestern Journal of Anthropology*, 7 (1951): 1-21; "Lexico-statistic dating of prehistoric ethnic contacts", in *Proceedings of the American Philosophical Society*, 96 (1952): 452-63; "Towards greater accuracy in lexicostatistical dating", in *IJAL*, 21 (1955): 121-37.

¹⁴See, for example, John A. Rea, "Concerning the Validity of Lexicostatistics", in *IJAL*, 24 (1958): 145-50; Knut Bergsland and Hans Vogt, "On the Validity of Glottochronology", in *Current Anthropology*, 3 (1962): 115-53; Saul Levin, "The Fallacy of a Universal List of Basic Vocabulary", in Horace G. Lunt, ed., *Proceedings of the Ninth*

ternational Congress of Linguists (The Hague: uton, 1964), 232-6.

¹⁵Swadesh, "Lexico-statistic Dating..."

¹⁶Swadesh, "Towards Greater Accuracy...", D. H. mes, "Lexicostatistics So Far", in *Current Anthropology*, 1 (1960): 3-44.

¹⁷G. L. Trager, in Bergsland and Vogt, "On the lidity of Glottochronology", *Current Anthropology*, (1962): 146.

¹⁸D. H. Hymes, "Lexicostatistics So Far."

¹⁹Thomas and Headley, "More on Mon-Khmer bgroupings", 411.

²⁰René Morizon, "Essai sur le dialecte des pulations Pears des Cardamones", Thèse complémen- ire pour le Doctorat ès-Lettres, (Paris: Les itions Internationales, 1936).

²¹Willlaim A. Smalley, *Outline of Khmu? ructure*. American Oriental Series Essays No. 2 altimore: American Oriental Society, 1961).

²²Thomas and Headley, *op. cit.*, 405-7.

²³*Ibid.*, 410.

²⁴Swadesh, "Towards Greater Accuracy..."

²⁵David Thomas pointed out, after giving nsiderable attention to my paper, that my figures pport the classification which appeared in Thomas d Headley 1970, with the exception of Alak, which ow places in North Bahnaric (see Gregerson, ith, and Thomas, "The Place of Bahnar within hnaric", in this volume).