



DEPT OF LINGUISTICS, PAYAP UNIVERSITY

August 2003

Karenic Language Relationships

A Lexical and Phonological Analysis

Ken Manson

TABLE OF CONTENTS

TABLE OF CONTENTS.....	I
LANGUAGE ABBREVIATIONS.....	II
1. INTRODUCTION.....	1
1.1 RESEARCH METHODOLOGY: AN OVERVIEW	2
1.2 PREVIOUS STUDIES	2
1.3 GENETIC TREES	4
2 LEXICAL RELATEDNESS ANALYSIS.....	7
2.1 LEXICAL SIMILARITY.....	7
2.2 THE RESULTS AND THEIR INTERPRETATION.....	8
3 PHONOLOGICAL RELATEDNESS ANALYSIS.....	11
3.1 DETERMINING THE NUMBER OF WORD FORMS TO COMPARE.....	11
3.2 CALCULATING THE DEGREE OF CHANGE	12
3.3 THE ALGORITHM	15
3.4 THE RESULTS	16
3.5 INTERPRETING THE RESULTS	17
4 IMPLICATIONS FOR KARENIC RESEARCH	19
5 CONCLUSIONS.....	20
APPENDIX 1: PERCENTAGE LEXICAL SIMILARITY.....	21
APPENDIX 2: PHONOLOGICAL SIMILARITY.....	22
APPENDIX 3: WORDLIST FOR PHONOLOGICAL ANALYSIS.....	23
BIBLIOGRAPHY	26

Language Abbreviations

Code	Location/Notes	Researcher	Source
Bwe	Bwe	Bennett	unpub. field notes
Bwe DE	Bwe, east of Daylo stream	Lar Baa	Lar Baa 2001
Bwe DW	Bwe, west of Daylo stream	Lar Baa	Lar Baa 2001
Bwe W	Blimaw Bwe	Luce	Luce 1985
Geba	Geba	Luce	Luce 1985
Geba1	Geba, “dialect 1”	Bennett	unpub. field notes
Geba2	Geba, “dialect 2”	Bennett	unpub. field notes
Gebah	Geba	Lar Baa	Lar Baa 2001
Geker	Geker	Lar Baa	Lar Baa 2001
Kayah E	East Kayah	Bennett	unpub. field notes
Kayah W	West Kayah	Bennett	unpub. field notes
Kayaw	Kayaw	Bennett	unpub. field notes
Manu	Manu	Bennett	unpub. field notes
Padaung	Padaung	Bennett	unpub. field notes
Paku	Luce’s Paku	Luce	Luce 1985
Paku K	Kathokhi Paku	Lar Baa	Lar Baa 2001
Paku S	Shokho Paku	Lar Baa	Lar Baa 2001
Palachi	Palachi	Lar Baa	Lar Baa 2001
Pa-O N	Shan States Pa’O	Bennett	unpub. field notes
Pa-O S	Pa’O Thaton (southern)	Bennett	unpub. field notes
Pwo D	Delta Pwo	Luce	Luce 1985
Pwo T	Tenasserim Pwo	Luce	Luce 1985
Sgaw	Sgaw	Luce	Luce 1985
Sgaw D	Sgaw Karen	Dai	Dai 1982
Sgaw K	Kya-In Sgaw	Lar Baa	Lar Baa 2001
Sgaw P	Papun Sgaw	Lar Baa	Lar Baa 2001
Sgaw R	Rangoon Sgaw	Lar Baa	Lar Baa 2001
Yinbaw	Yinbaw	Bennett	unpub. field notes
Yintale	Yintale	Bennett	unpub. field notes

1. Introduction

The Karenic languages are perceived by Tibeto-Burman linguists to form a unified block in comparison with other Tibeto-Burman languages (see for example Bradley 1997, Benedict 1972, LaPolla 2001, Matisoff 1991, Peiros 1998, Shafer 1973). However, very little work has been done in determining the internal relationships of the Karenic languages. And, moreover, this previous research has focused on the more populous Karenic languages: Sgaw, Pho and Pa'O, ignoring the numerous other Karen languages and dialects.

The Karen languages are found mostly in eastern Burma from southern Shan States southwards to the southernmost tip of Burma, as well as, along the western side of Thailand. Some Sgaw Karen have also migrated to the Andaman Islands. Bradley (1997:46) suggests a total population of 3.9 million, but notes that this is “substantially under enumerated”. The largest Karenic groups include Sgaw with 1.6 million; Pwo¹ with 1.4 million; Pa'O with 500,000 and Kayah with 250,000. There is somewhere between 6 and 10 million ethnic Karen, however, not all speak Karen languages. Many now speak only Burmese.

The present study provides an alternative, yet complementary, approach to the standard historical-comparative analysis.² The aims of the study include determining the lower-level clusters of Karenic language varieties and suggesting a possible internal relationship between these units; as well as providing a suitable research methodology that can be applied to a large number of wordlists of varying quality.

¹ The names Pho, Pwo and Phlong represent different speech varieties along the language continuum of Pwo.

² Even today after nearly a century of historical-comparative analysis the internal relationships between the different units which comprise Tibeto-Burman are still relatively unknown.

1.1 Research Methodology: An overview

Usually before any full-scale historical-comparative analysis is undertaken, a quicker lexical comparison is made to determine suitable candidate languages for the reconstruction. Sometimes this process is not formally carried out, but a quick inspection of the data is taken to see the gross similarities between languages. This approach of a quick inspection may introduce errors of omission especially if certain language varieties are excluded that have retained features from the proto-language.

Below I outline the research procedures undertaken in this paper.

First, all available wordlists for Karenic language varieties were collected and ranked according to their reliability on a scale from 1 (recorded by a modern linguist, i.e. collected in the last 30 years) to 5 (recorded by a government official, over 90 years ago) and on this basis wordlists were chosen for further analysis.³

An initial lexical relatedness analysis was done of 22 language varieties to find the relationships between these varieties. A phonostatistical analysis was then undertaken using the methodology similar to that of Baxter & Manaster Ramer (2000). The results of each approach are discussed at the end of the appropriate section. A final conclusion and summary is then provided.

1.2 Previous Studies

While the Karenic languages form a unified unit within Tibeto-Burman, very little modern linguistic analysis of the internal relationships of these languages has been undertaken.

Haudricourt (1946) reconstructed Proto-Karen on the basis of two literary languages: Sgaw and Pwo. He was able to describe and explain the tone splits that occurred within Karen using his earlier work on Vietnamese as a basis for showing the patterning of tonal variation across Sgaw and Pwo.

³ 85 wordlists have so far been collected. Many of them (21) are of Pwo speech varieties spoken in Northern Thailand.

Jones (1961) provides a strong, yet flawed, analysis of six speech varieties of Karenic languages.⁴ Burling's (1969) reanalysis provides more direction towards what the protoforms may have been. However, he also succumbs to the same error as Jones in not observing the distributional gaps associated with initial consonants and tones. Both appear unaware of the earlier work by Haudricourt (1946) and Luce (1959, see also 1985) who showed how tone variations in Karen are derivable from the phonation of the initial proto-consonant, which would have simplified their results. Furthermore, these two studies are based on only three clusters of languages, the so-called 'Central Karenic' languages were completely ignored, as Peiros states:

"The Karenic group includes numerous languages and dialects spoken in Central and South Burma, their exact number and pattern of relationships being still unknown. Jones (1961) discussed six Karenic dialects forming three subgroups, but perhaps a more detailed division is needed (Bradley 1996 [sic: 1997]). The phonological systems (especially the tones and vowels) of the Karen languages are rather complicated and the use of old dictionaries and word lists is therefore not straightforward. Any Proto Karenic reconstruction based mainly on Jones' lists will not fully represent the phonological system of the protolanguage (Burling 1969, Peiros 1989c). New dialectal data is needed (see, however, Ratanakul 1986).⁵

"Karenic dialect have undergone considerable phonological change, and their modern forms often do not resemble the protoforms. Thus it is very difficult to compare these dialects directly with other Sino-Tibetan languages. Without a Proto Karenic reconstruction one would come to various incorrect conclusions, such as that the Karenic group forms a very remote branch of the family." (Peiros 1998:179-80)

Kauffman (1993) describes "Central Karen" on the basis of three language clusters – Padaung, Geba and Kayah. He admits that the term 'central Karen' is a geographic label for those languages sandwiched between Pa'O to the north and Sgaw & Pwo to the south, although he notes that these languages do share some features that other Karen languages do not. However, he does not relate his results to those of Jones or Burling.

Weidert (1987) dedicates a chapter to the development of tonal systems in Karen languages, providing a number of etyma spread over the tonal categories, but again it is based on only three of the subgroupings, and does not provide any discussion on the relationships between the languages analysed.

⁴ In reality, only three languages were used: one language with three dialects (Moulmein Sgaw, Bassein Sgaw, and Palaychi – an aberrant dialect) and another with two dialects (Moulmein Pho, Bassein Pho), plus one other language (Taungthu = Pa'O).

⁵ However Ratanakul's work is an extended dictionary of Sgaw – another dialect of those already collected and analysed.

Solnit (1997) mentions the internal relationships of Kayah dialects and even briefer comments on more distant relationships.⁶

Bradley's (1997) summary of the present state of Tibeto-Burman includes a composite chart of relationships between the different Karenic language varieties based on personal communication with Jones, Lehman and Solnit – three linguists who have devoted significant research efforts towards Karen language relationships; however, no supporting evidence for the proposed relationships was given.

1.3 Genetic trees

Generally linguists have skirted the issue of specific internal relationships within Karen. However, three genetic diagrams for Karenic have been published.

I will not discuss the agnostic view that Karen has three branches – North (Pa'O), South (Sgaw and Pwo), and Central (the rest). This is a geographical division of the family for which no evidence has been presented. In fact, 1) Bennett (pc) refers to "Central" Karen as a wastebasket term for languages for which we still do not know the clear relationship; and 2) the greatest dialectal variation within Pa'O, a so-called North Karenic language, occurs in the southern area, not in its northern distribution.⁷

The first published diagram of Karenic language relationships was Jones (1961:83), see Figure 1. He argues for Pa'O and Pwo (his Pho) being more closely related to each other than Sgaw and Palaychi.

In Burling's reanalysis, he argues that "Pho and Sgaw seem to correspond to each other more consistently and with fewer complicating discrepancies than any of these correspond to either Palaychi or Taungthu" and that the "position of Taungthu appears even more extreme" (1969:4). This results in a diagram of Karenic language relationships as shown in Figure 2.

⁶ He is undertaking a major reconstruction of Karen but has not yet published any results of his research.

⁷ Furthermore, the Pa'O have a tradition of moving north to their present location in the last 2 centuries.

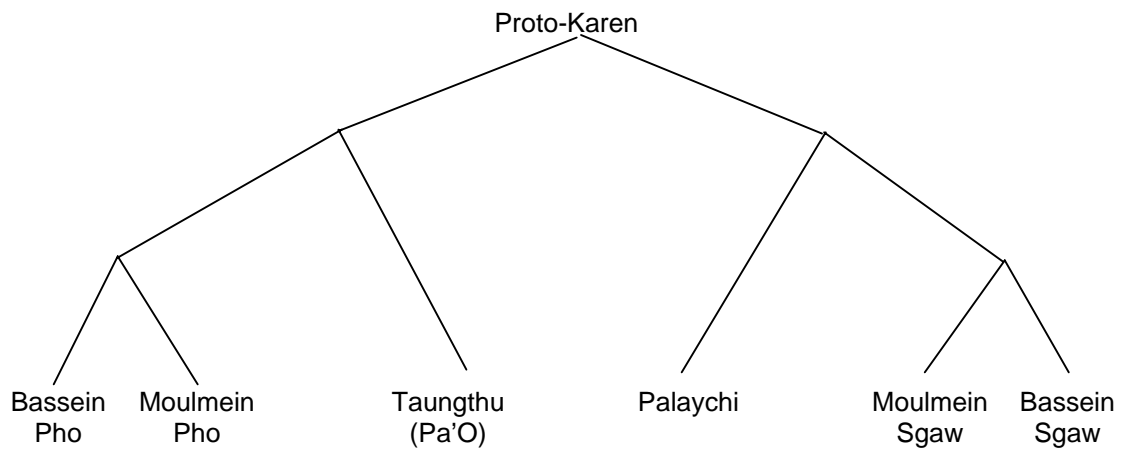


Figure 1: Karenic Language Relationships (after Jones 1961:83)

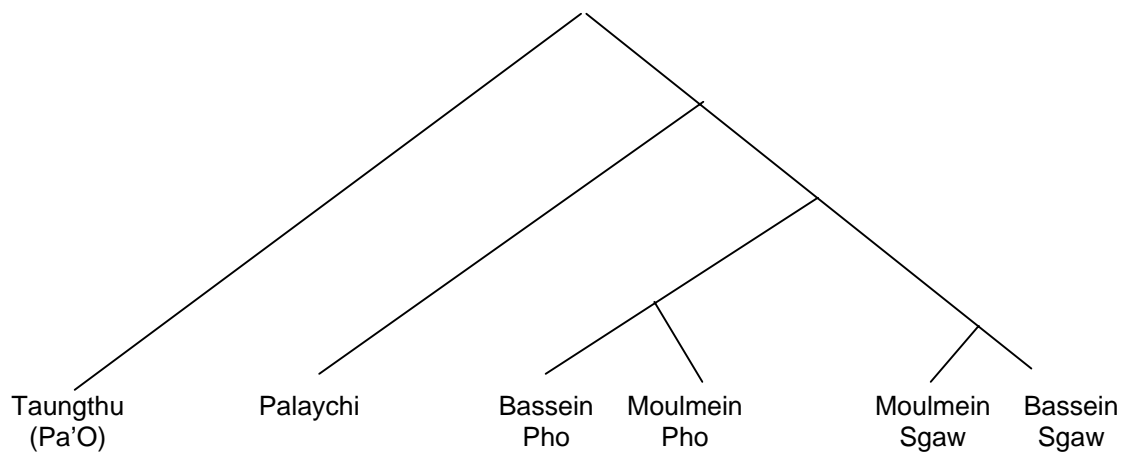


Figure 2: Karenic Language Relationships (after Burling 1969:4)

Kauffman (1993) also provides a suggested classification of Karenic languages, but again the “central” Karenic languages are defined in geographical terms, Figure 3.

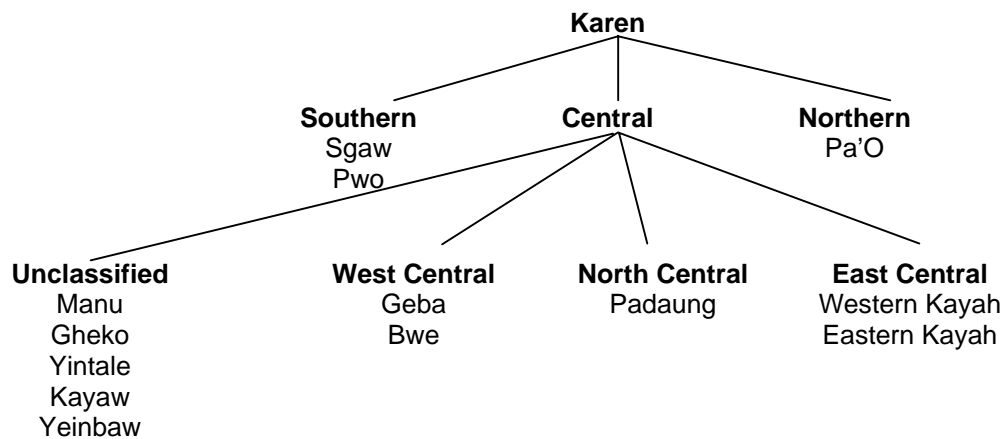


Figure 3: Karenic Language Relationships (after Kauffman 1993)

The final diagram to be discussed here is Bradley (1997). As noted above, Bradley does not provide any evidence for the classification presented. However, as it is the most complete classification of Karenic languages ever published it will be considered in this paper. It is shown in Figure 4.

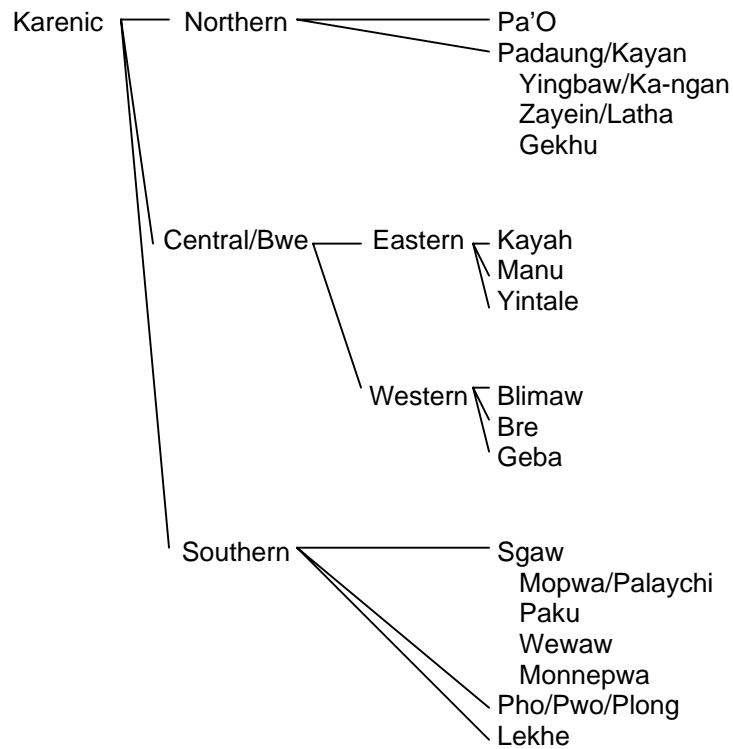


Figure 4: Karenic Language Relationships (after Bradley 1997:47)

2 Lexical Relatedness Analysis

An initial lexical relatedness analysis of 22 Karenic speech varieties was done to determine the internal relationship between these language varieties, and to assist in determining potential word meanings across languages that have been less likely to be replaced.

Wordlists collected by two researchers with a large spread of language varieties were chosen so as to minimise the variation caused by different methods/styles of recording the data. One hundred word meanings were chosen from the MSEA 406 wordlist; 89 of these also occur in the Swadesh 100 wordlist, and an additional 11 drawn from items more relevant to the South East Asian context (see Table 2). Each word meaning was compared across all language varieties to determine whether the forms were derived, or appeared to be derived from a common etymon or not, with each new variant form being marked by a different code.⁸ Once this was completed, a comparative score was calculated pair-wise for the language varieties to determine a percentage of apparent lexical similarity between them.

2.1 Lexical Similarity

The standard method used to create diagrams of lexical similarity (“trees”) in a lexicostatistic analysis is the ‘group average’ method, or more specifically, UPGMA ‘Unweighted Pair Group Method using arithmetic Averages’. Lohr (2000) provides a summary of this process in her analysis of Indo-European:

“we first group together the pair of languages with the highest shared cognacy score (here, Serbian and Russian). Next, the mean score of every other language’s pair wise cognacy scores with Serbian and Russian is found (i.e. the mean of English-Serbian and English-Russian, the mean of German-Serbian and German-Russian etc.) The matrix is redrawn with a ‘Slavic’ column and row, eliminating the Serbian and Russian columns and rows. Then, the current highest shared cognacy score in the matrix is found again, and the process is repeated, grouping language after language, until a complete classification is reached.” Lohr (2000:212)

The compiling and creation of a diagram that is of use for interpretation can be done by hand, or by using a number of available computer programs. This research used the NEIGHBOR,

⁸ *Prefixes* were included into the lexical similarity analysis. A word with *ke-* would be marked as different to a form that had a *te-* or *pe-* as its prefix. Vocalic prefixes were ignored. *Compounds* were marked as different to words which were composed of only one of the compound’s roots. *Tone* was ignored in the analysis, as tone is derived from the interaction of the initial proto-consonant and proto-rhyme of the syllable.

and KITCH programs from the PHYLIP package for biological classification and genetic analysis.⁹ This set of programs does the “donkey-work” of generating a diagram based on the percentages of dissimilarity between the language varieties.

The programs NEIGHBOR, and KITSCH were written to deal with data in the form of a matrix of pair-wise distances between all pairs of taxa (in this case, between the different languages).

The program NEIGHBOR can carry out the Neighbour-Joining method of Nei and Saitou (1987) and the UPGMA method of clustering. The Neighbour-Joining method constructs a tree by successive clustering the most similar pair of branches. The UPGMA method constructs a tree by clustering together pairs of branches using an average-linkage method of clustering. The program KITSCH can carry out the method of Fitch and Margoliash (1967) for fitting trees to distance matrices, as well as, the least squares method of Cavalli-Sforza and Edwards (1967).

Each of these four methods were applied to the data, producing four phenograms.

Appendix 1 lists the pair-wise apparent cognacy matrix for the 22 speech varieties used in this analysis.

2.2 The Results and their Interpretation

The results of the four methods (UPGMA, Neighbour-Joining, Fitch-Margoliash and Least-Squares) produced very similar phenograms. Figure 5 shows the UPGMA rooted phenogram for the language varieties analysed. From this figure we can see that the lexicostatistical algorithm has not been able to correctly group the language varieties. For example the six Bwe-Geba varieties are not linked directly together: Gebah, Bwe DE, and Bwe DW are separate from Bwe, Geba1, and Geba2.¹⁰ An alternative method for producing a phenogram from the data, the least-squares method, produces a topologically identical phenogram of relationships.

⁹ PHYLIP (Phylogeny Inference Package) version 3.5c, (Copyright 1999 Joseph Felsenstein & the University of Washington) is a set of programs designed for biological analysis of DNA sequences. It is also well suited for crunching the numbers derived from lexical and phonological comparisons. This package of programs is available (free of charge) from the author at <http://evolution.genetics.washington.edu/phylip.html>.

The reasons for this result is due to numerous influences. First, by keeping the number of researchers to a minimum, variations caused by different transcriptions were hoped to be restricted. However, sub-branches under the two primary branches (i.e. wordlists recorded by Lar Baa and Bennett) reflect the received structure for Karenic languages. Second, the language of elicitation was different for each researcher – this may skew the results towards the language of elicitation (eg. towards Sgaw in the case of Lar Baa). Third, there appear to be many cases of synonyms for closely related items. A more finely discriminating wordlist with clearly specified semantic items would reduce this problem. Fourth, and this is related to the previous reason, different dialects of each language may have chosen/preferred one synonym over another and so a lexical similarity analysis would consider these dialects to be more different than they actually are. Fifth, all languages have individuals who are more expert in language use than others, and if some informants were not as adept as others there would be skewing of the results.

The unclear result of the lexical similarity analysis is in agreement with Matisoff (2000), who argues that lexicostatistics (glottochronology) is an unsuitable method for subgrouping in Tibeto-Burman. The lexicostatistical algorithm is a rather crude procedure of comparison, especially when used for isolating monosyllabic-tending languages. A more refined algorithm is needed.

Thus, on the basis of a lexical similarity analysis we cannot, with any confidence, determine clear internal relationships between Karenic languages; in fact, it produces erroneous results.

¹⁰ However, there is a pattern observed in the results. When each language variety is marked by the researcher, we note that in Figure 5 from Bwe E to Palachi was recorded by one linguist and from Kayah W to Kayah by the other.

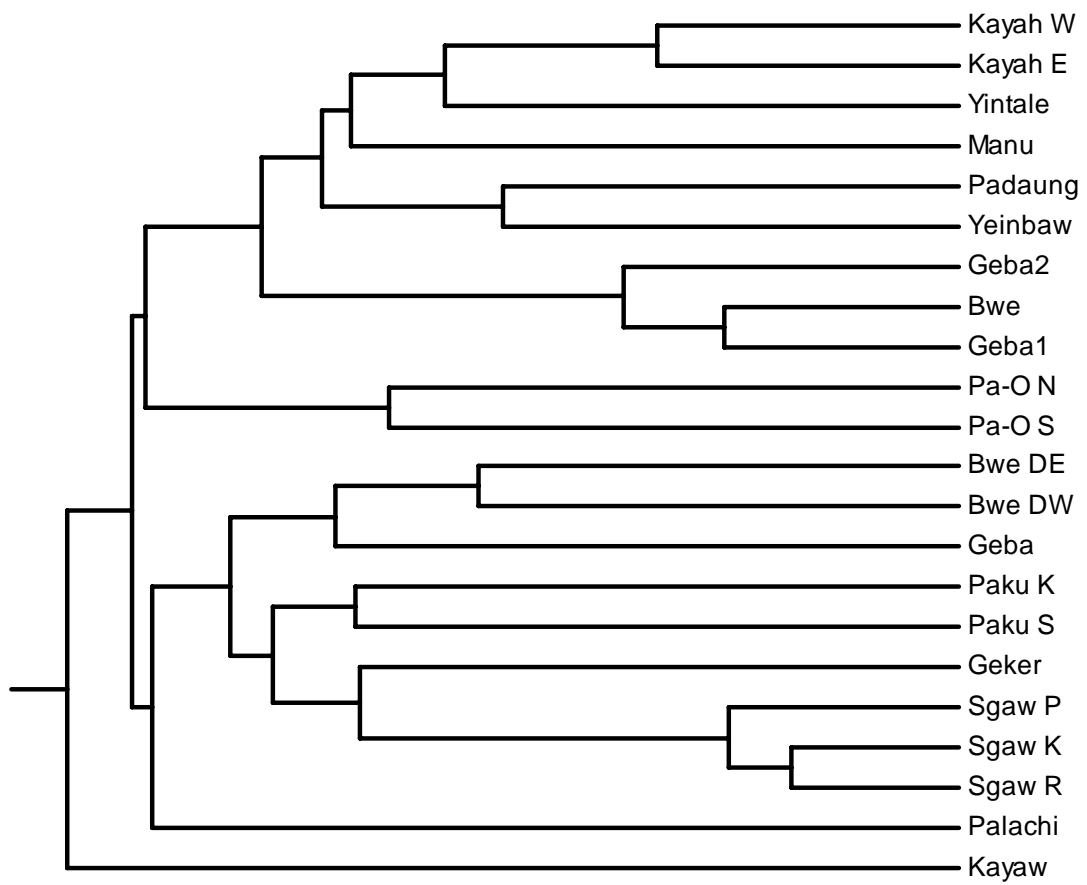


Figure 5: Karenic Lexical Similarity Phenogram

3 Phonological Relatedness Analysis

As lexical relatedness has been shown to be an unreliable methodology for subgrouping Karenic, a phonostatistical analysis was then applied to provide more sensitive criteria for determining the relationships between the language types. Due to the number of wordlists available for analysis, a smaller set of word meanings was selected.¹¹

3.1 Determining the number of word forms to compare

What is an adequate number of word-meanings to compare to be certain of genetic relationships? Is 100, 200, 436, 859, 866, or even 3000¹² enough? What if we want to say that we want to be more than 95% certain that our results reflect historical linguistic reality? How many words are truly needed to produce a result that we can have that confidence in?

Baxter & Manaster Ramer (2000) have shown that this number need not be very large to produce significant results. They took 33 word meanings derived ultimately from the Swadesh 100 that have been shown to be resistant to lexical change within Indo-European and then compared the English and Hindi forms. A very general comparison algorithm was used – the first sound of each word meaning was classified into one of the ten ‘Dolgopolsky Classes’¹³ The results showed that there were 9 matches between English and Hindi.¹⁴ And then based on a set of 1000 computer generated random trials, Baxter & Manaster Ramer showed that the probability of getting 9 or more matches was 11/1000, i.e. the probability that these 9 matches occurring by chance was 0.011. The Poisson probability formula used to approximate this distribution produced a similar result of 0.018 that there would be 9 or more matches out of 33 comparisons.

¹¹ There were in excess of 65 wordlists which were deemed to be of good quality. So for 65 language varieties, this would mean that there would be 2080 pair-wise comparisons and assuming only 100 words (from the complete 400+ words) in each list were analysed, and assuming 2 phonemes per root, there would be at least 416,000 comparisons to calculate (29 days at 2 seconds per comparison and working 8 hours per day!).

¹² See for example, the Swadesh 100; the Swadesh 200, Matisoff's CALSEA 200 (Note that Matisoff's 200 list is in fact a 211 word list – so does that mean a 5% error is acceptable?); SIL's Mainland Southeast Asia Wordlist 436 (or even their 406 list); Jones (1961); Bradley's Proto-Loloish list; the Academia Sinica 3000 wordlist.

¹³ These classes of sounds are those which are more likely to change over time into another member of the same class than into another class. For example, [p] – class 1 is more likely to change into [b] – class 1 than into [m] – class 5.

¹⁴ This algorithm produced some incorrect cognate matches and also ignored some true cognate matches.

Thus taking the lower probability as the limit to what may actually be random, Baxter & Manaster Ramer state that there is a 98% chance that these two languages are related, and a 2% chance that the two languages are not related.

So using a similar strategy to Baxter & Manaster Ramer, 34 word meanings were chosen for Karenic based on the initial 100 word-meaning of the lexical similarity comparison. These 34 word meanings were chosen so that all 8 Luce tone categories were evenly represented (see the next section for details).

3.2 Calculating the degree of change

Using the data from the lexical relatedness calculations, each word meaning was ranked by the number of variants observed. Initially the 100 word meanings were divided into four classes based on part-of-speech: nouns; numerals; verbs & adjectives; and determiners & question words to see whether there was any significant differences based on part-of-speech.¹⁵ Table 1 summarises the results:

Number	Nouns	Verbs / Adjectives	Number	Nouns	Verbs / Adjectives
1	3	-	9	2	-
2	5	-	10	3	3
3	5	3	11	2	3
4	5	4	12	1	4
5	6	3	13	3	2
6	8	7	14	1	-
7	5	3	15	-	1
8	9	3			

Table 1. Distribution of Variants by Part-of-Speech

The average number of variants for nouns was 6.3 with a standard deviation of 3.2, and for verbs the average was 7.8 ± 3.4 . The overall average for both categories was 6.9 ± 3.3 . On the basis of these results we cannot state that verbs or nouns as a class are more susceptible to change than the other. Figure 6 summarises the combined distribution of variants graphically.

¹⁵ As there was only two numerals and four determiners/question word-meanings in the list these were ignored in the analysis. Also the determiners & question words had an average of 15 variant forms per word.

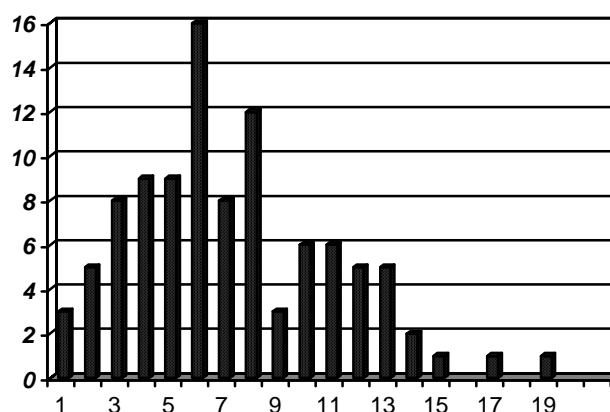


Figure 6. Frequency Distribution of Word-Meaning Variants

Figure 6 approximates a *Poisson density distribution*. We would expect that some words would undergo change/replacement more frequently than others, and even within a sample of words traditionally deemed to be core and, hence, less susceptible to replacement, there still is a great amount of variation. It is interesting to note that while most of these words are found in the Swadesh lists, there is a much greater than expected variation in the number of variants.¹⁶

Table 2 provides a listing of word-meanings by number of apparent variants.

As tone category is a dependent variable, predictable from the syllable shape and phonation of the initial consonant, word meanings were chosen so they were evenly distributed over all tone categories. Four word meanings were chosen from each tone category, with tone category 6 having six members, as this category constitutes 40% of all examples given in Luce (1985). Word meanings were chosen on the basis of low number of variants observed in the lexical relatedness analysis, and to have as great a diversity of phonological forms as possible (additional word meanings were needed to fill some tone classes as a number of forms from these tone classes had numerous variants), as can be seen in Table 3.

¹⁶ It is of note that for many Karenic groups they subdivide the language varieties on the basis of what the word for *what* and *who* are.

1	2	3	4	5	6	7	8
moon star water	sky monkey bone father fire	year leaf bamboo dog tongue eat kill bitter	sun iron skin blood mother weep die swim heavy	stone tree fish hair name ashes vomit sell deep	rain mountain root bite horn feather louse belly leg give shoot one big thick full new	thorn bird egg head flesh stand kick white	night earth tail ear hand knee person path smoke sit enter hot

9	10	11	12	13	14	15	17	19
chin heart two	mouth nail mat laugh many red	mist seed fly (v) see burn that	tooth liver dance short black	eye nose sing think all	wing this	green	where	who

Table 2. Word meanings by number of apparent variants

Luce TC	Gloss	Luce TC	Gloss	Luce TC	Gloss	Luce TC	Gloss
I	far	III	water	V	paddy rice	VII	son-in-law
I	name	III	die	V	bone	VII	brain
I	tongue	III	new	V	hand	VII	enter
I	person	III	black	V	spicy	VII	deep
II	silver	IV	father	VI ¹⁷	leaf	VIII	skin
II	white	IV	snake	VI	bitter	VIII	monkey
II	moon	IV	mother	VI	bamboo	VIII	pig
II	spear	IV	sun	VI	fire	VIII	dark
				VI	year		
				VI	star		

Table 3. Word-Meanings by Luce Tone Category

¹⁷ Tone Class VI has six representatives as this is the most numerous tone class observed from Luce's data (almost 40% of his etyma are from this class).

3.3 The Algorithm

Following Matisoff (1991) we assumed that Tibeto-Burman (and hence Karenic) was basically a monosyllabic language¹⁸ and so only the root syllable was considered in the comparison. Languages were then compared pair-wise by word meanings on a node-by-node basis with each node¹⁹ being assigned to one of three rankings, as can be seen in Table 4 (based on Blair 1990).

Then for each pair of languages a total was calculated by adding up the number of word-pairs that passed the limits set beforehand for determining sufficient similarity. These limits are shown in Table 5. So, for example, a word-pair with three nodes of comparison would need to have at least two Category 1 rankings and one Category 2 ranking to be considered cognate. Most word-pair comparisons involved two nodes. A percentage similarity was then calculated by dividing the number of apparent cognates with the total number of word-pairs compared. These results are summarised in Appendix 2.

¹⁸ Matisoff has also argued for a sesquisyllabic root, but it is still unclear whether the “prefixes” associated with roots involved any vocalic element (see LaPolla 2001). But for the Karenic languages these pre-root elements are almost always the same, and so have been ignored in the algorithm.

¹⁹ I use the term node to refer to the initial consonant element versus an optional medial consonant element versus the rhyme (i.e. the vocalic element plus optional final consonant). It is true that the glide element may reflect an earlier proto-vocalic element, but for the present study this is ignored. A reconstruction would need to consider this issue further.

Category 1	<ul style="list-style-type: none"> a. exact match b. rhyme differ by one feature c. phonetically similar segments observed more than 3 times in 100 word-pairs d. labialised versus non-labialised consonants e. palatalised versus non-palatalised consonants f. affricate versus fricative g. vowel that is intermediate between the ends of a diphthong h. voice difference word initial
Category 2	<ul style="list-style-type: none"> a. phonetically similar segments observed less than 2 times in 100 word-pairs b. rhyme differing by two or more features
Category 3	<ul style="list-style-type: none"> a. non-phonetically similar segments b. correspondence with zero in less than 3 out of 100 word-pairs
Ignored	<ul style="list-style-type: none"> a. reduced syllables and non-root syllables b. tone²⁰ c. voice quality (breathy/non-breathy) d. reduplicated syllables e. syllable final glottal stop

Table 4. Phonetic Similarity Algorithm

Number of Nodes	Category 1	Category 2	Category 3
2	2	0	0
3	2	1	0
4	2	1	1

Table 5: Assignment criteria for phonological similarity classes

3.4 The Results

Appendix 2 provides the pair-wise phonological similarity matrix for the 23 speech varieties used in the phonological similarity analysis. These 23 language varieties were chosen to provide a more extensive spread of language varieties than have been presented in previous analyses of Karenic language relationships.

As in the lexical similarity analysis, all four methods (UPGMA, Neighbour-Joining, Fitch-Margoliash and Least-Squares) were used to produce phenograms. All four methods clustered the languages into the same lower level clusters of languages. The main differences

²⁰ Voice quality (breathiness) and syllable final glottal stop are considered to be part of the tonal features of the syllable and so are ignored as the sample of word meanings was selected so as to cover all tone classes.

between these methods were in the higher level branches. The results derived from the UPGMA can be seen in Figure 7.

3.5 Interpreting the results

Figure 7 provides a tree diagram of Karenic languages. This figure correctly shows that language types that are most similar are clustered together. For example, the Bwe-Geba dialects are clustered together with each other rather than being spread over the phenogram as in the lexical relatedness analysis.

However, care must be taken in interpreting the diagram. This diagram does **not** show the genetic relationships between languages as would be produced by an in-depth comparative analysis. It is a product of the algorithm chosen for the analysis. A different algorithm may produce different results of the relationships.

The clustering of language varieties has absolute meaning, not just relative meaning. Looking at the phenogram shows seven clusters of languages – Sgaw/Paku; Pwo; Pa-O; Manu/Kayaw; Yeinbaw/Geker/Padaung; Bwe/Geba; and Kayah. Yintale appears to be a separate grouping, but further research is necessary to confirm its relationship with its near neighbours.

These seven clusters of languages are almost certainly significant groupings within Karen, but the higher-level relationships displayed in the figure are less likely to be an accurate reflection of reality. A reason for this higher-level uncertainty could include, for example, one group undergoing a recent phonological innovation – eg. vowel raising/diphthongisation.²¹

²¹ This appears to be the case for Kayah dialects. It should be noted also that some Pho dialects in northern Thailand have undergone similar phonological developments. These dialects will be included in further analysis.

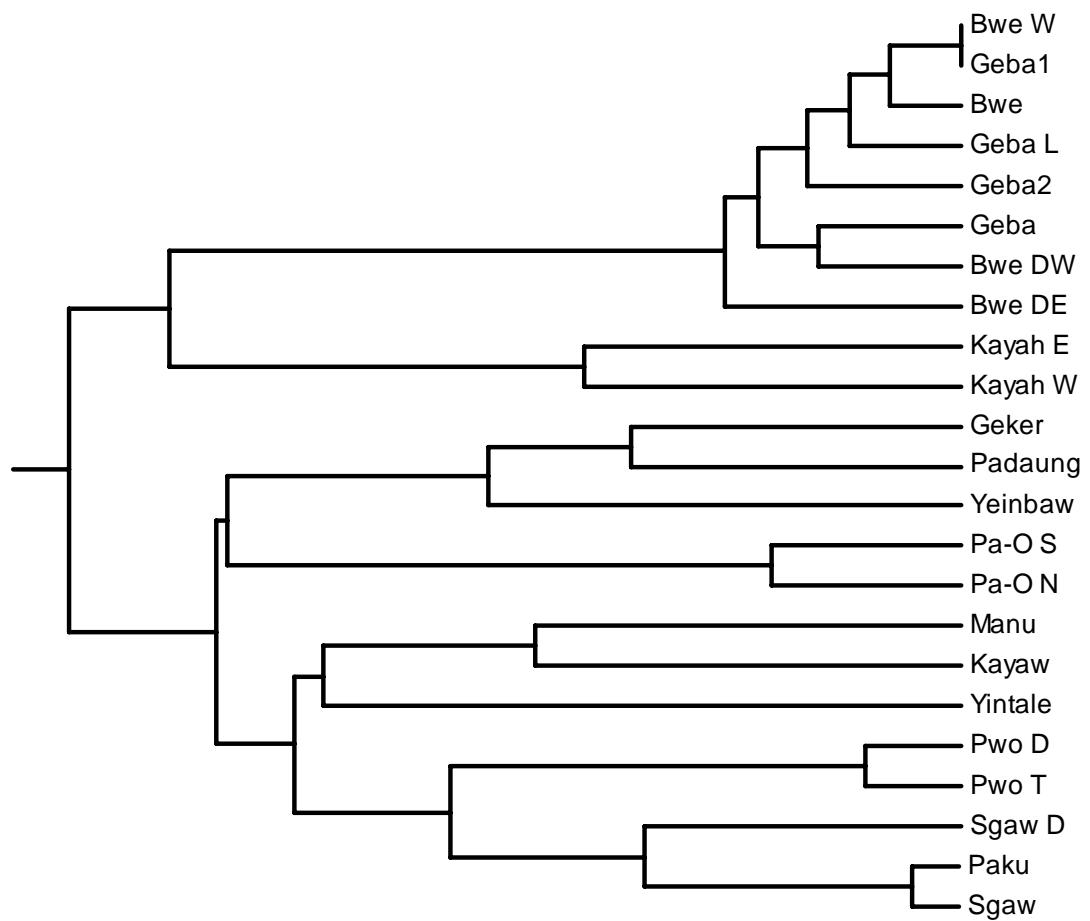


Figure 7. Phonological Similarity Phenogram

4 Implications for Karenic research

Almost all previous research of Proto-Karen has been based on data from the lower portion of Figure 7, namely the Sgaw, Pwo, and Pa-O languages (e.g. Haudricourt 1946; Jones 1961; Burling 1969), with Weidert (1987) being the exception by adding data from Bwe.

For a more definitive reconstruction of Proto-Karen there need to be more material collected and published, especially for the following clusters: Manu/Kayaw; Yeinbaw/Geker/Padaung; Bwe/Geba; and Kayah. Solnit's work on Kayah is of great importance, but needs to be added to. For example, Bennett (1991) provides data on two additional Kayah dialects.

Reconstructions based on only a few of the clusters of Karenic languages will produce results that will need to be revised as each new cluster of Karenic languages is added to the picture.

Looking back at the languages that previous researchers have considered, we see that the initial work by Haudricourt (1946) involves only two clusters, while the work of Jones (1961) and Burling (1969) is based on only three clusters, as is Kaufmann (1993). No linguistic study of Karenic languages has included the breadth necessary to develop an accurate and complete comparison.

It is interesting to compare the clusters determined by this research with the proposed relationships presented in Bradley (1997). Bradley presents three higher-level groupings (based on geography) with each then comprised of at least two subgroupings. These lower level subgroupings are matched by the clusters determined in this research, thus providing some quantitative backing to the proposed lower level genetic relationships.

5 Conclusions

The methodology used in this research can be transferred very easily to other language situations, finding its greatest application to those situations where good comparative research is still lacking and initial analysis needs to be done to direct further research.

This research provides a rapid appraisal of Karenic language relationships. This is the first step in determining more accurately the higher-level relationships between Karenic languages, as well as providing a set of key words to use in determining the position of new speech varieties (see Table 3).

The clusters of Karenic languages that have not yet received attention include the Padaung/Geker cluster (the Kayan), the Manu/Kayaw cluster (the Kayaw), and Yintale.

The Kayan cluster would be the next cluster for significant research as some work has already been done on Padaung spoken in the refugee camps of North Thailand, and the cluster involves a number of dissimilar speech varieties. It also retains final nasals and hasn't undergone the vowel raising of the Bwe/Geba and Kayah clusters.

Further research is also needed for a number of reported dialects including members of the Sgaw dialect cluster: Monnepwa and Mopwa, the Bwe/Geba language cluster; the Pa'O languages; and the Pwo languages.

Research is also needed regarding the sociolinguistic situation. There has been no published account of the Karen sociolinguistic situation since the notes in Lehman (1967).

Appendix 1: Percentage Lexical Similarity

Bwe DE	Bwe DW	Gebah	Geker	Paku K	Paku S	Palachi	Sgaw K	Sgaw P	Sgaw R	Kayah W	Kayah E	Kayaw	Manu	Bwe	Geba1	Geba2	Pa-O N	Pa-O S	Padaung	Yeinbaw	Yintale
.570	.404																				
.479																					
.333	.333	.294																			
.287	.313	.280	.353																		
.340	.394	.350	.373	.460																	
.287	.283	.190	.333	.240	.250																
.330	.384	.360	.431	.380	.450	.320															
.362	.364	.410	.471	.340	.430	.290	.770														
.383	.384	.390	.490	.360	.400	.310	.850	.820													
.223	.222	.180	.333	.250	.190	.180	.290	.250	.290												
.213	.172	.190	.294	.220	.170	.160	.210	.240	.210	.730											
.194	.153	.242	.314	.162	.172	.162	.162	.192	.141	.172	.273										
.250	.227	.235	.235	.296	.224	.173	.286	.235	.286	.510	.398	.206									
.340	.414	.330	.353	.280	.260	.210	.310	.270	.320	.360	.280	.222	.439								
.330	.374	.370	.353	.300	.260	.220	.350	.310	.350	.430	.310	.222	.449	.790							
.351	.333	.360	.353	.290	.240	.220	.290	.270	.330	.400	.290	.212	.459	.620	.780						
.223	.242	.190	.294	.220	.200	.190	.260	.290	.290	.340	.310	.212	.296	.260	.280	.290					
.223	.212	.160	.333	.210	.210	.200	.240	.270	.260	.220	.230	.222	.194	.220	.220	.210	.490				
.228	.227	.204	.451	.276	.224	.224	.296	.276	.327	.510	.418	.224	.423	.337	.398	.378	.357	.276			
.213	.202	.180	.216	.330	.230	.190	.310	.260	.300	.490	.360	.212	.367	.320	.400	.400	.350	.270	.592		
.255	.273	.210	.412	.270	.200	.160	.270	.220	.270	.590	.490	.192	.460	.360	.400	.390	.340	.250	.429	.450	

Appendix 2: Phonological Similarity

Kayah W	Kayah E	Kayaw	Manu	Bwe	Geba1	Geba2	Pa-O N	Pa-O S	Padaung	Yeinbaw	Yintale	Sgaw D	Pwo T	Pwo D	Sgaw	Paku	Bwe W	Geba	Bwe DE	Bwe DW	Geba	Geker
.765	.382	.235																				
.353	.265	.735																				
.559	.471	.529	.412																			
.529	.412	.559	.412	.971																		
.500	.412	.559	.382	.882	.941																	
.353	.235	.471	.412	.471	.441	.412																
.324	.265	.529	.471	.441	.441	.412	.882															
.353	.382	.529	.529	.500	.471	.412	.500	.588														
.353	.235	.500	.529	.500	.529	.529	.441	.559	.706													
.412	.412	.618	.588	.471	.441	.441	.529	.441	.471	.471												
.258	.226	.581	.516	.484	.516	.581	.645	.613	.548	.323	.613											
.412	.441	.647	.588	.588	.500	.471	.647	.647	.559	.588	.588	.677										
.412	.412	.529	.529	.412	.412	.412	.647	.647	.618	.500	.500	.613	.941									
.353	.353	.647	.559	.353	.382	.412	.471	.441	.500	.441	.618	.806	.735	.676								
.364	.333	.647	.576	.387	.364	.424	.485	.424	.545	.485	.636	.800	.727	.667	.970							
.618	.588	.618	.559	.941	1.000	.912	.471	.500	.529	.500	.500	.581	.529	.441	.471	.485						
.618	.500	.441	.471	.941	.912	.882	.471	.441	.500	.500	.412	.516	.471	.412	.441	.424	.941					
.515	.485	.455	.394	.788	.848	.879	.455	.455	.424	.455	.394	.633	.485	.424	.515	.500	.909	.848				
.529	.382	.529	.471	.853	.912	.882	.441	.441	.412	.471	.412	.548	.500	.382	.529	.485	.912	.853	.848			
.588	.412	.559	.441	.824	.853	.882	.441	.441	.441	.529	.412	.548	.471	.412	.471	.485	.882	.882	.848	.912		
.500	.382	.647	.618	.500	.471	.441	.559	.618	.794	.706	.559	.613	.618	.588	.559	.606	.500	.471	.455	.471	.529	

Appendix 3: Wordlist for Phonological Analysis

0	Luce TC	Gloss	Bwe DE	Bwe DW	Gebah	Geker	Paku K	Paku S	Palachi	Sgaw K	Sgaw P	Sgaw R	Kayah W
1	I	far	jiɬ	jiɬ	jiɬ	jiɬ	jɛiɬ	jiɬ	ziɬ	jiɬ	jiɬ	jiɬ	ziɛɬ
2	I	name	miɬ	miɬʔ	miɬ	miɬ	mɛiɬ	miɬ	ɱ	ɱiɬ	ɱiɬ	ɱiɬ	ɱɱiɬ
3	I	tongue	pliɬʔ	pleɬʔ	pliɬ	ple:iɬ	pyɛɬʔ	pyɛɬ	pliɬ	pleɬ	pleɬ	pleɬ	pliɬ
4	I	person	piɬjaɬ	bjaɬ	bjaɬ	praɬ	pwaɬ	pweɬ	plɔɬʔ	pyaɬ	pwaɬ	pwaɬ	prɛɬ
5	II	silver	xoɬ	xoɬ	xoɬ	ɬuɬ	ɬɛɬʔ	ɬɛɬʔ	soɬʔ	ɬɛɬ	seɬ	seɬ	ruɬ
6	II	white	boɬ	boɬ	boɬʔ	boɬ	pwiɬ	pweɬʔ	gwaɬʔ	waɬ	waɬ	waɬ	buɬ
7	II	moon	ɬeɬ	ɬeɬ	ɬeɬ	laɬ	liɬ	leɬʔ	laɬ	laɬ	laɬ	laɬ	leɬ
8	II	spear	baɬʔ	baɬ	baɬ	boɬ	baɬʔ	baɬʔ	boɬʔ	boɬ	boɬ	boɬ	bjaɬ
9	III	water	tʃiɬʔ	tʃiɬ	tʃiɬ	sʰɬuɬ	tʰɛiɬʔ	tʰiɬʔ	tʰiɬʔ	tʰiɬ	tʰiɬ	tʰiɬ	tʰieɬ
10	III	die	siɬ	tiɬ	tiɬ	ɬuɬ	sɛiɬ	θiɬ	tiɬʔ	tiɬ	tiɬ	tiɬ	ɛieɬ
11	III	new	ɬeɬ	ɬeɬ	ɬeɬ	ɬaɬ	saɬʔ	ɬaɬʔ	toɬʔ	ɬoɬ	ɬoɬ	ɬoɬ	ɛeɬ
12	III	black	piɬ	tiɬ	tiɬ	ɬɛɬʔ	sɛɬʔ	ɬuɬʔ	laɬʔ	tʰuɬ	tʰuɬ	tʰuɬ	lɛɬ
13	IV	father	paɬ	paɬʔ	paɬʔ	pʰaɬ	paɬ	paɬ	paɬ	pɛɬ	pɛɬ	pɛɬ	pʰeɬ
14	IV	snake	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ɬuɬ	ruɬ
15	IV	mother	moɬ	moɬʔ	muɬʔ	muɬ	muɬ	moɬ	muɬ	moɬ	moɬ	moɬ	muɬ
16	IV	sun	muɬʔ	muɬʔ	muɬ	mɛɬʔ	mɛɬ	mɛɬ	mɛɬ	muɬ	muɬ	muɬ	mɛɬ
17	V	paddy rice	buɬ	buɬʔ	buɬ	buɬ	buɬ	buɬ	buɬ	buɬ	buɬ	buɬ	buɬ
18	V	bone	kʰwiɬ	kʰwiɬ	kʰwiɬ	tʃweiɬʔ	xeiɬ	xwiɬ	xiɬ	xiɬ	xiɬ	xiɬ	kruiɬ
19	V	hand	ɬɛɬʔ	ɬɛɬʔ	suɬ	ɬɛɬ	ɬɛɬ	ɬɛɬ	ɬɛɬ	suɬ	suɬ	suɬ	kʰuɬ
20	V	spicy	heɬ	heɬ	heɬ	heɬ	ɬuɬ	ɬuɬ	heɬ	heɬ	heɬ	heɬ	hjaɬ
21	VI	leaf	ɬeɬʔ	ɬeɬ	ɬeɬʔ	laɬʔ	liɬʔ	leɬʔ	laɬ	laɬʔ	laɬʔ	laɬ	leɬ
22	VI	bitter	kʰeɬʔ	kʰeɬ	kʰeɬʔ	kʰaɬʔ	kʰeɬʔ	kʰeɬ	kʰaɬʔ	kʰaɬʔ	kʰaɬʔ	kʰaɬʔ	kʰeɬ
23	VI	bamboo	hoɬʔ	xoɬʔ	xoɬʔ	xwaɬʔ	wiɬ	weɬʔ	waɬʔ	waɬʔ	waɬʔ	waɬ	veɬ
24	VI	fire	meɬʔ	muɬ	miɬʔ	meiɬʔ	mɛɬ	mɛɬʔ	miɬʔ	meɬ	meɬʔ	meɬʔ	miɬ
25	VI	year	deɬʔ	diɬʔ	diɬʔ	nɛiɬ	niɬ	niɬʔ	nɛɬʔ	niɬʔ	niɬʔ	niɬ	naɬ
26	VI	star	sʰeɬʔ	sʰeɬʔ	sʰeɬ	sʰaɬʔ	tʃhiɬʔ	tʃhiɬʔ	sʰaɬ	tʃhiɬʔ	sʰaɬʔ	sʰaɬ	sʰeɬ
27	VII	son-in-law	maɬʔ	maɬ	maɬʔ	maɬ	maɬ	maɬʔ	-	maɬʔ	maɬʔ	maɬʔ	mɛɬ
28	VII	brain	nuɬʔ	nuɬ	nuɬʔ	noɬ	noɬ	nuɬʔ	nuɬʔ	nuɬʔ	nuɬʔ	nɛɬʔ	nɛɬ
29	VII	enter	-	nuɬʔ	nɛɬʔ	nuɬ	noɬ	nuɬ	-	nuɬ	nuɬʔ	nuɬʔ	nuɬ
30	VII	deep	joɬʔ	joɬʔ	joɬʔ	ɬɛɬʔ	-	-	joɬ	joɬʔ	joɬʔ	joɬʔ	ɛaɬ
31	VIII	skin	pʰeɬ	pʰeɬʔ	pʰɛiɬʔ	pʰeɬʔ	boɬ	boɬ	beɬʔ	boɬʔ	pʰiɬʔ	pʰɛiɬʔ	pʰaɬ
32	VIII	monkey	joɬ	joɬ	joɬ	joɬʔ	zoɬ	ɬuɬʔ	zuɬ	ɬuɬʔ	ɬuɬʔ	ɬuɬʔ	ɛaɬ
33	VIII	pig	tʰɔɬ	tʰɔɬʔ	tʰoɬʔ	tʰɔɬʔ	tʰoɬ	ɬoɬʔ	tʰoɬʔ	tʰoɬʔ	tʰoɬʔ	tʰoɬʔ	tʰjaɬ
34	VIII	dark	tʃiɬ	kʰiɬʔ	tʃhiɬʔ	kʰuɬʔ	kʰɛiɬʔ	tʃhiɬ	sʰiɬʔ	kʰiɬʔ	kʰiɬʔ	kʰɛiɬʔ	kʰiɬ

0	Luce TP	Gloss	Kayah E	Kayaw	Manu	Bwe	Geba1	Geba2	Pa-O N	Pa-O S	Padaung
1	I	far	jaJ	jeɬ	ʒiJ	jɛJ	jɛJ	ʒɪɬ	ŋjaɣ	ŋjaɬ	ɟuɣJ
2	I	name	mɯiJ	mɪɬ	mɪJ	mɪɬ	mɪɬ	mɪɬ	minɬ	mɪnɬ	mjɯũJ
3	I	tongue	pliJ	pleiɬ	plɛJ	plɛɬ	plɛɬ	plɛɬ	pʰreɬ	pʰreɬ	pleiJ
4	I	person	pʰreJ	jɔɬ	jɔJ	pwɛɬ	pwɛJ	pjaJ	louɬ	lɔuɬ	praJ
5	II	silver	ruɬ	rouɬ	βuJ	hoɬ	hoɬ	hoɬ	rʌnɬ	rʌmJ	ʔwauũɬ
6	II	white	buɬ	buɬ	boɬ	boɬ	boɬ	boɬ	bwaɬ	bwaJ	buɬ
7	II	moon	leɬ	laɬ	laJ	ʔɛɬ	hɛɬ	leɬ	laɬ	laJ	laɬ
8	II	spear	bɛɬ	boɬ	boJ	baɬ	baɬ	baɬ	baŋɬ	baŋJ	bouũɬ
9	III	water	tʰʌɬ	tʰɪɬ	ʃʰiJ	tʰiɬ	tʰiɬ	tɕʰiɬ	tʰiJ	tʰiJ	ʃʰuɬ
10	III	die	tɕʌɬ	θiɬ	siJ	θiɬ	θiɬ	θiɬ	siJ	θɪJ	suɬ
11	III	new	tɕeɬ	θaɬ	saJ	θɛɬ	θɛɬ	θɛɬ	saJ	θaɬ	suɬ
12	III	black	loJ	θuɬ	lɔɔJ	tiɬ	tiɬ	θiɬ	pʰrenJ	pʰrenɬ	lɔuJ
13	IV	father	pʰeɬ	paiɬ	pʰaJ	paɬ	paɣ	paɬ	pʰaɣ	pʰaɣ	pʰaɬ
14	IV	snake	ruJ	rouJ	βoɬ	wuɬ	ɭuɬ	wuɬ	ruɣ	ruɣ	Rɯuɬ
15	IV	mother	mɔJ	mɯɪJ	mɯɔɬ	moɬ	mɔɬ	mɔɬ	mɣɬ	mɯɣ	mɯɣuɬ
16	IV	sun	mɔJ	mɯJ	mɔɬ	mɯɬ	mɯɬ	mɯɬ	mɯɣ	mɯɣ	mɯɣuɣ
17	V	paddy rice	boɬ	bɯɬ	buɬ	buɬ	buɬ	buɬ	buɬ	bɯɣ	buuɬ
18	V	bone	kʰɪɬ	ʃʰuɬ	kʰwiɬ	kʰwiɬ	kʰwiɬ	kʰwiɬ	cutJ	sʰwaɬ	ʃʰwiJ
19	V	hand	kʰuɬ	cɯJ	kʰɔɬ	tɕuɬ	suɬ	cɯɬ	cɯɬ	suJ	cɯJ
20	V	spicy	hɛɬ	hɛɬ	hɛɬ	hɛɬ	hɛɬ	hɛɬ	hatɬ	hapɬ	haiɬ
21	VI	leaf	leJ	laJ	laJ	leɬ	ʔɛɬ	leɬ	laɬ	laɬ	laJ
22	VI	bitter	kʰeJ	kʰaJ	kʰaɬ	kʰɛɬ	kʰɛɬ	kʰɛɬ	kʰaɬ	kʰaɬ	kʰaJ
23	VI	bamboo	veJ	ʌoJ	waɬ	hoɬ	hoɬ	hoɬ	waɬ	waɬ	ʌaJ
24	VI	fire	meJ	mɪJ	miɬ	miɬ	hmiɬ	miɬ	meɬ	mɛɬ	mɛiJ
25	VI	year	naJ	dɪJ	neɬ	deɬ	deɬ	deɬ	neŋɬ	neiɬ	neiJ
26	VI	star	tɕʰeJ	ʃʰaJ	sʰaɣ	ʃʰɛɬ	ʃʰɛɬ	sʰɛɬ	cʰaɬ	sʰaɬ	sʰaJ
27	VII	son-in-law	meɬ	mɔɬ	mɔɬ	mɔɬ	mɔɬ	mɔɬ	makɬ	maɬʔ	mɔɬ
28	VII	brain	nɔɬ	nɯɬ	nɔɔJ	nɔɬ	nɯɬ	nɯɬ	nɔkɬ	naɣɬʔ	nɔJ
29	VII	enter	noɬ	nɯɬ	nɯɬ	nɯɬ	nɯɬ	leɬ	niɬʔ	niɬʔ	nɯJ
30	VII	deep	jeɬ	tʃuɬ	ʒuɬ	jɔɬ	jɔJ	jɔJ	ʒɔɬʔ	jɔɬʔ	cɯu
31	VIII	skin	pʰaɬ	pʰɛɬ	beɬ	pʰɛɬ	pʰɛɬ	pʰɛɬ	piɬ	pʰɛɬʔ	pʰɛɬ
32	VIII	monkey	jɔɬ	ʒuɬ	jɔɔJ	jɔɬ	jɔɬ	ʒuɬ	jouɬʔ	jɔɬʔ	jɔɬ
33	VIII	pig	tʰɛJ	tʰoɬ	tʰoɬ	tʰɔɬ	tʰɔɬ	tʰɔɬ	tʰɔɬʔ	tʰɔɬʔ	tʰauJ
34	VIII	dark	kʰɪɬ	kʰiɬ	kʰiɬ	kʰiɬ	kʰiɬ	kʰiɬ	kʰɛJʔ	kʰɛJʔ	kʰuɬ

0	Luce TP	Gloss	Yeinbaw	Pwo D	Yintale	Sgaw D	Pwo T	Sgaw	Paku	Bwe W	Geba
1	I	far	zɛ˧˥	jaɪ̃˧	zɿ˧	ji˧	jai˧	ji˧	ji˧	ji˧	ji˧
2	I	name	mju̯ɿ˧	mɛ̃˧	mwaj˧	mi˧	mɛ̃˧	mi˧	mi˧	mɿ̃˧	mɿ̃˧
3	I	tongue	plɛ˧	pʰle˧	plj˧	ble˧	pʰle˧	ple˧	ble˧	blɿ˧	bli˧
4	I	person	prə˧	pə˧ pa˧	pjə̃˧	bua˧	hə˧	pra˧	bra˧	bwe˧	bwe˧
5	II	silver	rwo˧	se˧	mə˧	tsɛ˧	se˧	se˧	-	hu˧ ru˧	ho˧
6	II	white	bo˧	ɬwa˧	bu˧	wə˧	wa˧	wa˧	wa˧	ɬu˧	bu˧
7	II	moon	la˧	la˧	la˧	la˧	la˧	la˧	la˧	le˧	hle˧
8	II	spear	bā˧	pʰā˧	cɔ˧	--	pʰā˧	ɬo˧	ɬo˧	ɬa˧	ba˧
9	III	water	ʃʰu˧	tʰi˧	tʰai˧	tʰi˧	tʰi˧	tʰi˧	tʰi˧	cʰi˧	tʰi˧
10	III	die	θu˧	θi˧	sai˧	tθi˧	θi˧	θi˧	θi˧	θi˧	θi˧
11	III	new	θi˧	θā˧	sa˧	tθɔ˧	θā˧	θo˧	θo˧	θɛ˧	θɛ˧
12	III	black	lɔu˧	θɔ̃˧	sū˧	tθu˧	θɔ̃˧	θu˧	θu˧	θi˧	θi˧
13	IV	father	pʰa˧	pʰa˧	pʰai˧	ba˧	pʰa˧	pa˧	pa˧	pa˧	pa˧
14	IV	snake	rɔu˧	Rʷu˧ ru˧	ru˧	tθɛ˧	RU˧	Rɿ˧	Rɿ˧ ru˧	ru˧	wu˧
15	IV	mother	mu̯ɿ˧	mo˧	zɿ˧	mo˧	mo˧	mo˧	mu˧	mo˧	mo˧
16	IV	sun	mɔ˧	mɿ˧	mɿ˧	mu˧	mɿ˧	mɿ˧	mɿ˧	mu˧	mū˧
17	V	paddy rice	hu˧	ɬɿ˧	bɔ˧	bɔ˧	ɬɿ˧	ɬɿ˧	ɬɿ˧	ɬu˧	bu˧
18	V	bone	ʃʰwe˧	hwɪ˧	kʰrʷai˧	xu̯i˧	χwi˧	χRi˧	χRi˧	kʰwi˧	kʰwi˧
19	V	hand	cə˧	su˧	cə˧	--	su˧	sɿ˧	sɿ˧	tcu˧	su˧
20	V	spicy	hai˧	rai˧	hai˧	hɛ˧	re˧	he˧	he˧	he˧	he˧
21	VI	leaf	la˧	la˧	la˧	la˧	la˧	la˧	la˧	le˧	hle˧
22	VI	bitter	kʰa˧	kʰa˧	kʰa˧	kʰa˧	kʰa˧	kʰa˧	kʰa˧	kʰɛ˧	kʰɛ˧
23	VI	bamboo	ma˧	wa˧	va˧	wə˧	wa˧	wa˧	wa˧	hu˧	hu˧
24	VI	fire	mi˧	me˧	mi˧	mɛ˧	me˧	me˧	me˧	mi˧	hmi˧
25	VI	year	nei˧	nɛ̃˧	nai˧	nj˧	nɛ̃˧	ni˧	ni˧	de˧	de˧
26	VI	star	sʰa˧	ʃa˧	sʰa˧	tsʰa˧	ʃa˧	sʰa˧	sʰa˧	ʃe˧	sʰɛ˧
27	VII	son-in-law	mə˧	ma˧?	dai˧	ma˧	ma˧?	ma˧?	ma˧	ma˧	ma˧
28	VII	brain	nɔ̃˧	no˧?	nu˧	nu˧	nau˧?	nu˧?	nu˧	nu˧	no˧
29	VII	enter	nɔ̃˧	nə˧?	nu˧	ni˧	nau˧?	nɿ˧?	nɿ˧ nu˧	nu˧	ni˧
30	VII	deep	zau˧	ju˧?	zɿ˧?	ju˧	ju˧?	ju˧?	ju˧	ju˧	ju˧
31	VIII	skin	pʰe˧	pʰæ˧?	bui˧	bɛ˧	pʰai˧?	pʰi˧?	pʰi˧	pʰe˧	pʰe˧
32	VIII	monkey	zəu˧	o˧?	zu˧	u˧	au˧?	ɿ˧?	u˧	jo˧	jo˧
33	VIII	pig	tʰau˧	tʰu˧?	tʰo˧	tʰo˧	tʰu˧?	tʰu˧?	tʰu˧ tʰo˧	tʰo˧	tʰo˧
34	VIII	dark	kʰə˧	kʰe˧?	kʰi˧	--	kʰai˧?	kʰi˧?	kʰji˧	kʰi˧	kʰi˧

Bibliography

- Baxter, William H. & Alexis Manaster Ramer. 2000. *Beyond lumping & splitting: probabilistic issues in historical linguistics*. vol 1:167-188 in Colin Renfrew, et. al. eds.
- Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.
- Bennett, J. Fraser. (n.d.). *field notes – Karenic languages*.
- Bennett, J. Fraser. 1991. *Two More Kayah Li Dialects: A Comparison of daw te ma and daw sho pla dialects with Western and Eastern Kayah*. Paper presented at 24th ICSTLL, Bangkok
- Blair, Frank. 1990. *Survey on a shoestring: a manual for small-scale language survey*. Dallas: SIL.
- Bradley, David. 1997. *Tibeto-Burman languages and classification*. p1-71 in David Bradley ed. *Papers in Southeast Asian linguistics No 14: Tibeto-Burman languages of the Himalayas*. Pacific Linguistics A-86.
- Burling, Robbins. 1969. *Proto-Karen: a reanalysis*. Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics vol 1.
- Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. *Phylogenetic analysis: models and estimation procedures*. *Evolution* 32: 550-570.
- Dai Qingxia, Huang Bufan, Xu Shouchun, Chen Jiaying, Wang Huiyin eds. 1982. *A Tibeto-Burman lexicon*. Beijing: Central Institute of Nationalities Publishing House.
- Dolgopolsky, Aaron B. 1986. *A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia*. p27-50 in Vitalij V. Shevoroshkin & T. L. Markey eds. & trans. *Typology relationship and time*. Ann Arbor: Karoma.
- Fitch, W. M., and E. Margoliash. 1967. *Construction of phylogenetic trees*. *Science* 155: 279-284.
- Haudricourt, André-Georges. 1946. *Restitution du Karen Commun*. *Bulletin de la Société Linguistique de Paris* 42.1.103-111.
- Jones, Robert B. Jr. 1961. *Karen linguistic studies*. Berkeley: University of California Press.
- Kauffman, William. 1993. *The great tone split and Central Karen*. MA Thesis, University of North Dakota, Grand Forks.
- LaPolla, Randy J. 2001. *Overview of Sino-Tibetan morphosyntax*. ms.
- Lar Baa. 2001. *The phonological basis of a northwest Karenic orthography*. MA Thesis: Payap University, Thailand.
- Lehman, F. K. 1967. *Kayah society as a function of the Shan-Burma-Karen context*. p1-104 in J. H. Steward ed. *Contemporary change in traditional societies*. Urbana: University of Illinois Press.
- Lohr, Marisa. 2000. *New approaches to lexicostatistics and glottochronology*. vol 1:209-222 in Renfrew, et. al. eds.
- Luce, G. H. 1959. *Introduction to the comparative study of Karen languages*. *Journal of the Burma Research Society* 42:1-18.
- Luce, G. H. 1985. *Phases of Pre-Pagan Burma: languages and history*. 2vols. Oxford: Oxford University Press.
- Matisoff, James A. 1973. *Tonogenesis in Southeast Asia*. in Larry Hyman ed. *Consonant types and tones*. California: University of South California.
- Matisoff, James A. 1991. *Sino-Tibetan linguistics: present state and future prospects*. *Annual Review of Anthropology* 20:469-504.
- Matisoff, James A. 2000. *On the uselessness of glottochronology for the subgrouping of Tibeto-Burman*. vol 2:333-371 in Renfrew, et. al. eds.

- Nei, M. and W.-H. Li. 1979. *Mathematical model for studying genetic variation in terms of restriction endonucleases*. Proceedings of the National Academy of Sciences, USA 76: 5269-5273.
- Peiros, Ilia. 1998. *Comparative linguistics in Southeast Asia*. Pacific Linguistics Series C-142.
- Renfrew, Colin, April McMahon & Larry Trask. eds. 2000. *Time depth in historical linguistics*. Cambridge: The McDonald Institute for Archaeological Research. (2 volumes).
- Shafer, Robert. 1973. *Introduction to Sino-Tibetan*. Wiesbaden: Otto Harrassowitz.
- Simons, Gary. 1977. *Phonostatistic methods*. p155-186 in Richard Loving ed. *Language variation and survey techniques*. Ukarumpa: SIL (Workpapers in New Guinea languages vol 21).
- Simons, Gary. 1979. *Language variation and limits to communication*. Technical Report No. 3. Ithaca, NY: Department of Modern Languages and Linguistics, Cornell University.
- Solnit, David. 1997. *Eastern Kayah Li: grammar, texts and glossary*. Hawai'i: University of Hawai'i Press.
- Suriya Ratanakul. 1986. *Thai – Sgaw Karen Dictionary*. Bangkok: Mahidol University.
- Weidert, Alfons K. 1987. *Tibeto-Burman tonology: a comparative account*. Amsterdam: John Benjamins.