spoken south of Sop Moei were significantly different from their variety. The lexicostatistic analysis, which showed the two varieties to be 90% lexically similar, further indicates that these two varieties are less similar than most of the other varieties in this study .

Respondents from LEKHO reported that speakers in all the villages in their area spoke exactly the same dialect.

4. A Comparison of the Pwo Karen Dialects of Northern Thailand

In previous sections we have compared various varieties of Pwo Karen within the context of the geographical, i.e., provincial, boundaries in which they are located. While this is useful for assessing differences among the speech types that are found in each province, it nevertheless represents artificial divisions. In this section we will consider how the eighteen varieties investigated in this study interrelate and will mention some conclusions that can be made from a comparison of their lexical and phonological similarities.

4.1 Lexical Similarities

In considering lexical similarities among a fairly large group of speech varieties, the analyst is faced with a number of dimensions along which a comparison can be made. It is difficult to avoid skewing the results when the corpus is so large.

The goal of our study is to identify dialect groupings among the varieties of Pwo Karen spoken in northern Thailand, and to determine which varieties are typical of each grouping. To accomplish this goal we will need to quantify the degree of lexical similarity that each variety displays to all the other varieties in question. In an effort to strengthen the analysis we will look at the data from several perspectives.

In Table 4 below, percentages of lexical similarity for all of the speech varieties in the corpus, pair by pair, are presented in a matrix.

BAAN DOI (Chiang Rai)
96 WANG LUANG (Chiang Mai)
96 96 MAE LEK (Lamphun)
96 96 96 SALOK (Phrae)
95 95 97 95 PA LAU (Lamphun)
95 97 95 94 93 TIAN ANG (Chiang Mai)
97 95 95 95 95 94 NONG RAK (Lamphun)
94 94 95 96 96 93 94 MAE JONG FAI (Phrae)
97 95 95 94 94 93 95 93 THUNG KHONG (Chiang Rai)
94 95 93 93 92 96 92 92 92 YANG KHROK (Chiang Mai)

94 92 93 93 93 92 92 95 91 92 BAAN KLANG (Lampang)
93 92 91 92 92 92 91 93 91 91 97 RUAM MIT (Chiang Rai)
93 92 92 92 92 91 92 94 92 90 96 97 MAE SAAN (Lampang)
93 93 94 92 93 92 93 93 92 91 90 89 90 HUAI LA (Lamphun)
93 94 93 91 91 93 92 91 92 91 90 89 90 94 HUAI MUANG (Chiang Mai)
92 94 91 91 90 94 91 91 90 93 91 91 90 90 89 MAE KHONG KLANG (Chiang Mai)
93 94 93 92 91 96 91 91 91 94 90 90 89 90 91 92 THUNG PHRAW (Mae Hong Son)
90 91 88 88 88 91 88 87 89 93 86 86 85 88 88 90 90 LEKHO (Mae Hong Son)

TABLE 4   Lexical Similarity Percentages in the Northern Provinces[14]

There are a number of observations we can make from Table 4 regarding the
relationships among the varieties of Pwo Karen in northern Thailand. The most obvious
is that most of the varieties spoken in four of the northern provinces, Chiang Rai,
Chiang Mai, Lamphun and Phrae, are very closely related (See Map 3 below). This fact
is illustrated by the relatively high lexical similarity of the first four locations in the
matrix: Baan Doi, Wang Luang, Mae Lek and Salok. These four varieties are the four
varieties that are lexically most similar to the other members of the corpus (See Table 5
below). I would like to propose that we call the group of dialects to which these
varieties belong NORTHERN PWO KAREN.

---

[14]In order to allow for a second type of lexical comparison I have divided Table 4 into two sections. This allows us
to compare each variety to a smaller corpus (the upper group of ten) in addition to the entire eighteen varieties. The
ten varieties in the first section are considered to be more typical of the Pwo Karen varieties in northern Thailand.
Although the lexical similarity percentages do not significantly change between Yang Khrok and Baan Klang, the
division is not completely arbitrary since Pwo Karen in northern Thailand generally consider Baan Klang, Ruam Mit
and Mae Saan to be much more divergent in nature than the varieties in the upper set.
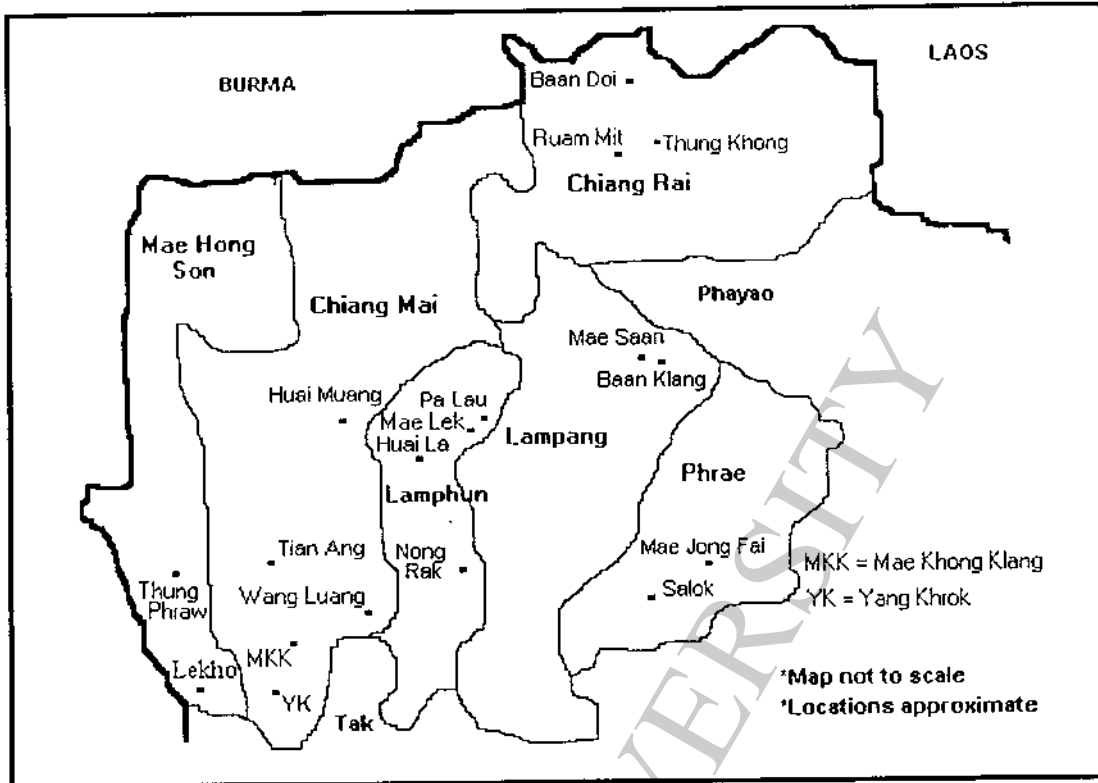
MAP 3 Word List Locations

4.1.1 The Lampang Varieties

We have seen that in Lampang province there are two dialects of Pwo Karen. Although there are some differences between them, they are nevertheless very closely related with 97% lexical similarity. However, when we compare them with the other speech types spoken in the North we find that lexical similarity between the Lampang varieties and other northern varieties is lower than the average lexical similarity percentages between the more typical[15] northern varieties, i.e., those listed in the upper set of Table 4. Whereas the upper ten varieties of the dialect show 94.6% lexical similarity with one another, the Lampang varieties show only 92.3% lexical similarity with the upper ten varieties.

One variety that is closely related to the Lampang variety is found in RUAM MIT

15In this case "more typical" is being defined as those varieties of Pwo Karen that are more typical of the entire dialect area.

(Muang District of Chiang Rai). In fact, RUAM MIT shows 97% lexical similarity with both of the Lampang varieties. This fact, along with the phonostatistic analysis to be presented in Section 4.4, and the clear historical link between the Pwo Karen in Lampang and the Pwo Karen in Chiang Rai Muang District, shows evidence that we have a very similar speech type represented in these two areas. I would like to propose that we call the group of dialects to which these varieties belong LAMPANG PWO KAREN. This speech type is spoken in the four Pwo Karen villages in Lampang province and by the Pwo Karen in the three mixed villages of Muang District in Chiang Rai province. It is 92.3% lexically similar to NORTHERN PWO KAREN (the upper set in Table 4).

4.1.2 The Huai La Variety of Pwo Karen

As was noted in the section on Lamphun province, the HUAI LA (Baan Hong) dialect is significantly different from the other speech varieties of the province being an average of 93.6% lexically similar with the other three varieties, whereas the latter are an average of 95.7% lexically similar with one another.

Like the speech varieties of LAMPANG PWO KAREN, the variety of Pwo Karen spoken in HUAI LA, in northern Lamphun Province, diverges significantly from the "typical" speech varieties of the North, being an average of 92.6% lexically similar with NORTHERN PWO KAREN. In its lexicon, HUAI LA is most closely related to MAE LEK, another Lamphun variety, and HUAI MUANG, a variety spoken in Chiang Mai, being 94% lexically similar with each of these.

Historically, however, HUAI LA is related to the HUAI MUANG variety in a way that the lexicostatistics do not reveal. Both of these villages are the result of a migration from southern Mae Hong Son province. Due to their close contact with Northern Thai, speakers in HUAI LA have adopted a large number of loan words that are not present in the HUAI MUANG data. Where HUAI LA often agrees with MAE LEK and the other Lamphun varieties in using a loan word, HUAI MUANG typically retains the Pwo Karen

term, making it more different than it probably once was. If these loan words were excluded, the similarity between these two varieties would be much more evident.

The relationship of HUAI LA and HUAI MUANG to LEKHO in southern Mae Hong Son is not very strong at only 88% cognate. It is possible that this significant difference is due to dialect change that has taken place since the HUAI LA and HUAI MUANG speakers migrated from Mae Hong Son. Alternatively, it is possible that the two villages represented a significantly different speech type from the LEKHO variety even prior to their migration. Further research in Mae Hong Son and Tak provinces may help to clarify the relationships between these varieties.

4.1.3 The Chiang Mai Varieties of Pwo Karen

In the course of interviews in the four easternmost provinces, although comprehension among speakers in those provinces was always reported at 100%, most speakers identified Chiang Mai province as the area where communication among Pwo Karen speakers begins to break down. Most respondents from the four provinces reported that they could understand most of what the Chiang Mai Pwo Karen say if they speak slowly. Others reported difficulty in understanding Chiang Mai speakers even under the best conditions.

Several further observations can be made from the data presented in Table 4. First, it is apparent that the Pwo Karen of Chiang Mai province do not speak a homogeneous speech form. There are significant differences that we have noted already. However, what we have not pointed out is that some varieties of Pwo Karen in Chiang Mai are very close to the Pwo Karen of the other three provinces, where NORTHERN PWO KAREN is spoken, whereas others are more divergent. WANG LUANG is fairly typical of NORTHERN PWO KAREN showing an average of 95.4% lexical similarity with the varieties in the upper half of Table 4. TIAN ANG is also fairly typical with an average of 94.4% lexical similarity. YANG KHROK is less typical with 93.2% lexical

similarity and HUAI MUANG and MAE KHONG KLANG are less typical yet showing an average of 92.1% and 91.7% lexical similarity respectively with the varieties in the upper half of Table 4.

4.1.4 The Mae Hong Son Varieties of Pwo Karen

Finally, although more data will be necessary to come to any definitive conclusions, there are a couple of observations that can be made concerning the relationships between the Mae Hong Son varieties of Pwo Karen and the other northern varieties. Table 4 indicates that THUNG PHRAW is lexically very close to the variety spoken in TIAN ANG (96% lexical similarity), and is also close to that spoken in WANG LUANG and YANG KHROK (94% lexical similarity). Its relationship to the other varieties of NORTHERN PWO KAREN is less strong (about 92.6% lexical similarity). LEKHO, on the other hand, is most similar to the YANG KHROK variety (93% lexical similarity), followed by WANG LUANG and TIAN ANG (91% lexical similarity). It is only 89.3% lexically similar to NORTHERN PWO KAREN.

4.2 Two Perspectives on Lexical Similarity

In this type of study, if there is a greater number of speech varieties from a certain geographical area, the figures that indicate overall lexical similarity will tend to be higher for those that are from that area since it is better represented in the corpus. To help counteract this and to move towards a conclusion as to which varieties are the most typical representatives of NORTHERN PWO KAREN, we can consider how each variety compares with the ten most "typical" varieties as well as how each compares with the entire corpus. Table 5 illustrates the change in rankings using the two systems.

Rank Against Entire Corpus	Village Name	Lexical Similarity Percentage	Rank Against "Top 10"	Lexical Similarity Percentage
1	BAAN DOI	94.44	1	95.5
2	WANG LUANG	94.37	2	95.4
3	MAE LEK	94.06	3	95.3
4	TIAN ANG	93.87	7	94.4
5	SALOK	93.62	4	95.0
6	NONG RAK	93.44	5	94.7
7	MAE JONG FAI	93.44	8	94.2
8	PA LAU	93.37	6	94.6
9	THUNG K.	92.94	9	94.0
10	YANG KHROK	92.56	10	93.2
11	BAAN KLANG·	92.56	11	92.7
12	MAE SAAN	92.00	15	92.0
13	RUAM MIT	91.94	16	91.8
14	HUAI LA	91.81	12	92.6
15	THUNG P.	91.75	13	92.6
16	HUAI MUANG	91.50	14	92.1
17	MAE K. K.	91.25	17	91.7
18	LEKHO	89.30	18	88.6

TABLE 5 Overall Rating versus Top Ten Rating

The differences in lexical similarity between each variety and the entire corpus compared with the similarity between each variety and the "top ten" are not extensive. BAAN DOI, WANG LUANG and MAE LEK retain the top three rankings. There are only two significant changes of rank in the "top ten" varieties: PA LAU and TIAN ANG. When we consider the ranking of TIAN ANG in relation to the entire corpus, we find that by having more varieties from Chiang Mai in the corpus, TIAN ANG appears more typical than it probably actually is. If we simply compare the two varieties which have the highest average lexical similarity to the entire corpus from each of the provinces where NORTHERN PWO KAREN is spoken, TIAN ANG would be ranked even lower in terms of average lexical similarity, i.e., seventh out of eight as opposed to seventh out of ten.

The elevation of PA LAU in the ranking is the converse of what has happened to TIAN ANG. When the varieties with lower averages for lexical similarity are excluded,

especially the Chiang Mai and Mae Hong Son varieties, PA LAU's position is elevated, i.e., when PA LAU is compared with the typical NORTHERN PWO KAREN varieties as opposed to the entire corpus it is seen to be more "typical" of NORTHERN PWO KAREN than was previously thought.

4.3 The Significance of the Statistical Measure

Finally, we turn to the question of how the figures presented in Tables 4 and 5 should be interpreted. Most linguists would agree that dialectal conclusions based on an absolute value for lexical similarity (or apparent cognate) percentages are inappropriate. A better method is to determine what range is significant between the various percentages. This will give us an idea of dialect groupings. To do this I will follow the techniques described by Simons in Loving and Simons (1977:75ff.).

Simons uses two factors to determine significant differences in apparent cognate percentages: size of data corpus and confidence level. The former factor in our case would be the size of the word list, or more precisely the number of words from the word list that were compared for any given pair of speech varieties.[16] The second factor, confidence level, represents a more subjective component. The confidence factor quantifies the analyst's evaluation of the reliability of the data. An average research situation, where the researcher has good bilingual language helpers, is assigned a confidence level of .10. If the data is the result of a "short intensive period of field work" it is assigned a higher confidence level: .20. Finally, if it is the result of many years of field work in the language it is given a confidence level of .30.

Simons provides a number of tables that indicate significant differences between apparent cognate percentages for corpi of various sizes (50, 100, 150, 200, 300, and

---

[16]Loan words and words that were later seen to have resulted from miscommunication in the elicitation process were excluded. An example of the latter is when we tried to elicit the Pwo Karen word for 'blood,' using Northern Thai, and were given the word for 'choose.' The Thai words for these two concepts differ only in the point of articulation of the final stop (which is unreleased in both cases).

500 words) with confidence levels of .01, .05, .10, .20 and .30. For the present study the confidence level is considered to be between .10 and .20, and closer to the latter than the former. Of the 405 words that were elicited at each location, between 384 and 395 were compared for any given pair of speech varieties. In Table 6 below, I have extracted the relevant figures from Simons' tables. Since the size of the corpus is between the two largest sizes that he presents, I have included the figures for both of these corpus sizes as well as for confidence levels of both .10 and .20.[17]

Corpus Size	300 Words		500 Words	
Confidence Level	.10	.20	.10	.20
	89-93	89-92	89-92	89-91
	90-93	90-92	90-93	90-92
	91-94	91-93	91-94	91-93
	92-95	92-94	92-95	92-94
	93-96	93-95	93-95	93-95
	94-97	94-96	94-96	94-96
	95-98	95-97	95-97	95-97
	96-99	96-98	96-98	96-98

TABLE 6 Significant Differences between Cognate Percentages

Table 6 lists sets of percentages for the two confidence levels. If a group of percentages fall within the range of one of these sets at the appropriate confidence level then they are considered to not be "significantly different." For example, if we compared three language varieties that were 89%, 90% and 93% lexically similar (at a confidence level of .10), we could say that the pairs showing 89% and 93% lexical similarity have essentially the same statistical relationship as the pairs showing 89% and 90% lexical similarity even though the former pairs are separated by four percentage points as opposed to one percentage point with the latter pairs.

17The four tables, with two sizes of word lists and two confidence values, will also show that the analysis presented below would be essentially the same regardless of which of the four sets of figures is used.

07870

In contrast, a given pair of speech varieties showing a percentage of lexical similarity that falls outside of the 89%-93% range would be considered to be significantly different, i.e., that pair of speech varieties would be viewed as having a different relationship to one another than the relationship that the varieties falling within that range had among themselves.

Using Table 6, we can evaluate the relative significance of the differences between various lexical similarity percentages in Table 4. For example, if we compare the percentage of lexical similarity between TIAN ANG and PA LAU with those between TIAN ANG and WANG LUANG we find that the former pair has a lexical similarity percentage of 93% and the latter pair shows 97% lexical similarity. Is this four percentage point difference significant? To answer this we can refer to Table 6 above. We will refer to the figures for a 300-word corpus with a confidence level of .20. Looking at the table we find ranges of 93-95%, 94-96% and 95-97%. There is no statistical difference between pairs of figures which fall within any of these ranges. However, 93% and 97% do not fall within any of these three ranges of percentages. Thus, we can say that the relationship between TIAN ANG and PA LAU is statistically significantly different to the relationship between TIAN ANG and WANG LUANG.

In our study, we are interested in the overall relationship of each variety to all of the other varieties. In order to determine this relationship based on significant differences we can use another method that Simons has outlined. The data in the corpus are arranged into groups based on an internal and an external criterion. "Internally the significance group can contain no percentages that are significantly different. Externally, any significance group must be significantly different from all other significance groups" (p. 83). In order to adjust the matrix to reflect significant differences the percentages in the matrix are replaced with the average percentage of the significance group of which they are a member (See Table 7).

Baan Doi (Chiang Rai)
94 Wang Luang (Chiang Mai)
94 94 Mae Lek (Lamphun)
94 94 94 Salok (Phrae)
94 94 94 94 Pa Lau (Lamphun)
94 94 94 94 94 Tian Ang (Chiang Mai)
94 94 94 94 94 94 Nong Rak (Lamphun)
94 94 94 94 94 94 94 Mae Jong Fai (Phrae)
94 94 94 94 94 94 94 94 Thung Khong (Chiang Rai)
94 94 94 94 91 94 91 91 91 Yang Khrok (Chiang Mai)

94 91 94 94 94 91 91 94 91 91 Baan Klang (Lampang)
94 91 91 91 91 91 91 94 91 91 94 Ruam Mit (Chiang Rai)
94 91 91 91 91 91 91 94 91 91 94 94 Mae Saan (Lampang)
94 94 94 91 94 91 94 94 91 91 91 91 Huai La (Lamphun)
94 94 94 91 91 94 91 91 91 91 91 91 91 94 Huai Muang (Chiang Mai)
91 94 91 91 91 94 91 91 91 94 91 91 91 91 91 Mae Khlong Klang (Chiang Mai)
94 94 94 91 91 94 91 91 91 94 91 91 91 91 91 91 Thung Phraw (Mae Hong Son)
91 91 87 87 87 91 87 87 87 94 87 87 87 87 87 91 91 Lekho (Mae Hong Son)

TABLE 7 Lexical Similarity Percentages Based on Significant Differences

We can now calculate each speech variety's closeness to a postulated
prototypical speech type, basing the calculation on the number of varieties that are not
significantly different from it. Once again we will look at both the figures that give the
lexical similarity percentages among all of the varieties, and those that give the lexical
similarity percentages between each variety and the ten most typical varieties. The
higher the number in either case the *less* typical the variety.
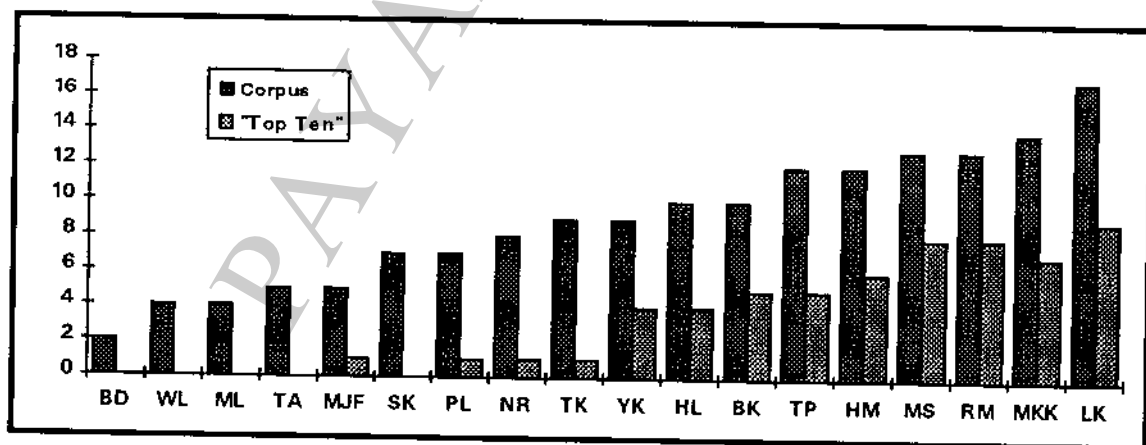
FIGURE 1 A Summary of Significant Differences

Figure 1 helps to clarify and strengthen some of the conclusions we have reached thus far. BAAN DOI has the highest degree of lexical similarity to both the entire corpus and the "top ten." It is significantly different from only two varieties of the overall corpus (MAE KHONG KLANG and LEKHO) and is not significantly different from any of the ten most typical varieties. In fact, the three most typical varieties remain the most typical using this system of comparison. It could be argued that WANG LUANG is the most typical variety of NORTHERN PWO KAREN since the only varieties it differs from are the three varieties of the LAMPANG PWO KAREN and LEKHO.[18]

TIAN ANG, which is significantly different from five varieties of the corpus but from none of the ten most typical varieties is substantiated as a typical variety of NORTHERN PWO KAREN. MAE LEK, MAE JONG FAI, SALOK, PA LAU, NONG RAK and THUNG KHONG all continue to also appear to be typical varieties of NORTHERN PWO KAREN.

YANG KHROK, on the other hand, appears even less likely to be a typical variety of this dialect grouping. It is significantly different from nine varieties in the corpus, four of which are included among the nine most typical varieties. Of these four, two are from Lamphun (PA LAU and NONG RAK), one from Phrae (MAE JONG FAI) and one from Chiang Rai (THUNG KHONG). I will reserve judgment on YANG KHROK until we have considered its phonological relationship to the rest of the corpus.

4.4 Phonological Similarities

The phonostatistic analysis presented in Table 9 below provides a statistical measure of the relative phonological distance between the various speech varieties. The figures represent ratios among the various speech varieties and show the average phonetic divergence between corresponding phones within lexically similar items. Every pair of corresponding phones has been assigned a numerical value representing

---

[18]It is interesting to note that judging from the lexicon only, it appears that if someone were interested in promoting literature among the Pwo Karen of northern Thailand using the Baan Doi variety, it is likely that the materials would be usable by speakers of Lampang Pwo Karen as well Northern Pwo Karen.

the degree of phonetic difference (DD) between the two phones. When comparing two speech varieties, the values representing phonetic variation are added and then divided by the number of total correspondences to yield the ratios in Table 9. The lower the figure, the closer the relationship between the speech varieties. In very general terms, we can say that speech varieties in this study having a phonological relationship lower than .30 degrees of difference per phone are "close" phonologically, while numbers higher than that indicate significant phonological divergence between pairs of speech varieties.

BAAN DOI (Chiang Rai)
21 WANG LUANG (Chiang Mai)
24 22 MAE LEK (Lamphun)
21 21 16 SALOK (Phrae)
21 25 16 16 PA LAU (Lamphun)
21 10 23 22 26 TIAN ANG (Chiang Mai)
23 29 22 19 18 28 NONG RAK (Lamphun)
25 24 26 21 25 24 29 MAE JONG FAI (Phrae)
09 21 22 19 18 21 21 25 THUNG KHONG (Chiang Rai)
24 13 25 23 27 10 29 25 23 YANG KHROK (Chiang Mai)
28 25 26 22 26 28 30 24 27 27 BAAN KLANG (Lampang)
35 29 29 30 32 30 33 32 35 30 32 HUAI LA (Lamphun)
37 33 36 35 35 34 39 36 36 35 26 38 RUAM MIT (Chiang Rai)
36 33 33 29 33 33 34 33 35 32 23 36 21 MAE SAAN (Lampang)
27 21 28 27 30 22 33 28 28 23 29 25 32 34 HUAI MUANG (Chiang Mai)
26 14 25 24 26 12 29 24 24 10 27 31 33 31 24 MAE KHONG KLANG (Chiang Mai)
25 14 25 24 28 15 31 26 24 17 27 32 33 32 24 17 THUNG PHRAW (Mae Hong Son)
27 17 26 26 30 18 34 30 25 19 29 35 36 37 27 20 15 LEKHO (Mae Hong Son)

TABLE 8 Phonological Degrees of Difference Ratios[19]

Some general comments can be made concerning the phonological relationships within provincial boundaries based on the figures presented in Table 8.

---

[19]The figures in Table 8 have been multiplied by 100 to make the table more readable.

4.4.1 Lamphun Province

In our discussion of the lexical relationships within Lamphun province we noted that the MAE LEK and PA LAU varieties should probably be viewed as a single dialect. This claim is strengthened by their phonological relationship. In Table 9, we see that MAE LEK and PA LAU have only .16 degrees of difference per phone. They are very close phonologically.

We can also note that HUAI LA is more divergent phonologically than the other Lamphun varieties. It has .29 DDs per phone with MAE LEK and .32 DDs per phone with PA LAU. This difference is important as we consider the relationship of the HUAI LA variety to the typical NORTHERN PWO KAREN varieties.

4.4.2 Phrae Province

As we have seen above, the two varieties of Pwo Karen spoken in Phrae province are quite close lexically, having 96% lexical similarity. They are also close phonologically since SALOK and MAE JONG FAI have only .21 DDs per phone. Table 9 also helps to clarify the relationship between the Phrae and Lamphun varieties of Pwo Karen. In our discussion of lexicostatistics we saw that the SALOK variety of Phrae province shows 96% lexical similarity with MAE LEK and 95% with PA LAU. Phonologically, SALOK has only .16 DDs with these two varieties. This is less phonological difference than between SALOK and MAE JONG FAI. Thus, the Wang Chin (SALOK) dialect of Phrae and the Mae Thaa dialect (MAE LEK and PA LAU) of Lamphun are very close both in terms of lexicon and phonology.

4.4.3 Chiang Rai Province

In our discussion of the lexicostatistic analysis of the Chiang Rai varieties it was noted that there are two different dialect groupings represented in this province. The Pwo Karen dialects spoken in BAAN DOI and THUNG KHONG belong to one of these

groupings (NORTHERN PWO KAREN) while the variety spoken in RUAM MIT belongs to the other grouping (LAMPANG PWO KAREN). These claims are strengthened by the phonostatistic analysis presented above.

The BAAN DOI and THUNG KHONG varieties have the strongest phonological relationship of any pair of varieties in this study. They have only .09 DDs per phone. This is further evidence that they should be regarded as a single dialect of NORTHERN PWO KAREN. RUAM MIT's phonological relationship to the other Chiang Rai varieties is even more divergent than its lexical relationship. It has .37 DDs per phone with BAAN DOI and .36 DDs per phone with THUNG KHONG. It is phonologically closest to MAE SAAN, one of the other varieties of LAMPANG PWO KAREN, having only .21 DDs per phone.

4.4.4 Lampang Province

In the lexicostatistical analysis we concluded that among the varieties of Pwo Karen spoken in northern Thailand there are (at least) two dialect groupings: NORTHERN PWO KAREN and LAMPANG PWO KAREN. The two varieties spoken in Lampang province and the RUAM MIT variety spoken in Chiang Rai are significantly different lexically from the other northern varieties. They are also different phonologically. As we have seen, the RUAM MIT variety is very different phonologically from even its near geographical neighbors, but it is close to the MAE SAAN variety in Lampang province and also close to the BAAN KLANG variety (.26 DDs). MAE SAAN and BAAN KLANG are also close phonologically having .23 DDs per phone. However, their phonological relationship to the NORTHERN PWO KAREN varieties is much weaker with BAAN KLANG having .26 DDs per phone on the average and MAE SAAN having .33 DDs per phone on the average. RUAM MIT has .35 DDs per phone with the NORTHERN PWO KAREN varieties.

4.4.5 Chiang Mai Province

As we compare the results of the lexicostatistical analysis of the Chiang Mai speech varieties with the phonostatistical analysis of the same varieties, some interesting observations emerge. We noted that WANG LUANG, TIAN ANG and YANG KHROK are very closely related lexically. HUAI MUANG and MAE KHONG KLANG, on the other hand, are more divergent, with both of the latter being most closely related to WANG LUANG in their lexicon. We find that WANG LUANG and TIAN ANG are also very closely related phonologically. They have only .10 DDs per phone. TIAN ANG and YANG KHROK also have only .10 DDs per phone, while WANG LUANG and YANG KHROK have just .13 DDs per phone. Phonologically, MAE KHONG KLANG is also very closely related to WANG LUANG (.14 DDs), TIAN ANG (.12 DDs), and YANG KHROK (.10 DDs). HUAI MUANG, on the other hand is more divergent. It has .21 DDs per phone with WANG LUANG, .22 DDs per phone with TIAN ANG, and .23 DDs per phone with YANG KHROK. While these latter figures do not represent a significant difference when compared with the overall corpus, they do indicate a weaker relationship than the relationships among the other Chiang Mai varieties.

4.4.6 Mae Hong Son Province

Finally, we will look at the phonological relationship among the Mae Hong Son varieties and between these two varieties and the rest of the corpus. The first thing that is apparent from Table 9 is that the Mae Hong Son varieties are not nearly as different from the other varieties phonologically as they are lexically. This is particularly true in the case of LEKHO. Although LEKHO has a comparatively low 89.3% lexical similarity with the other varieties, it shows a fair degree of phonological similarity with .265 DDs per phone on the average. If the varieties of LAMPANG PWO KAREN are not included, this figure is .247 DDs. While these figures do not indicate a strong phonological relationship, they do indicate that LEKHO is much closer to the other varieties

36

phonologically than it is lexically. Although LEKHO is only 90.6% lexically similar to the Chiang Mai varieties, it has only .20 DDs per phone on average with them. It is closest phonologically to THUNG PHRAW, the other variety from Mae Hong Son, with just .17 DDs per phone.

As in the lexicostatistical analysis, THUNG PHRAW is closest to the Chiang Mai varieties, in particular the TIAN ANG variety. It has 96% lexical similarity with TIAN ANG and just .14 DDs per phone. It averages 93.4% lexical similarity with the Chiang Mai varieties and just .17 DDs per phone.

4.5 Clarifying the Phonological Relationships

In order to further develop the significance of the data presented in Table 9 we will consider the phonological relationship of each speech variety to all of the other varieties in this study. Figure 2 illustrates the relative phonological similarity of each speech variety to all the other varieties in degrees of difference.
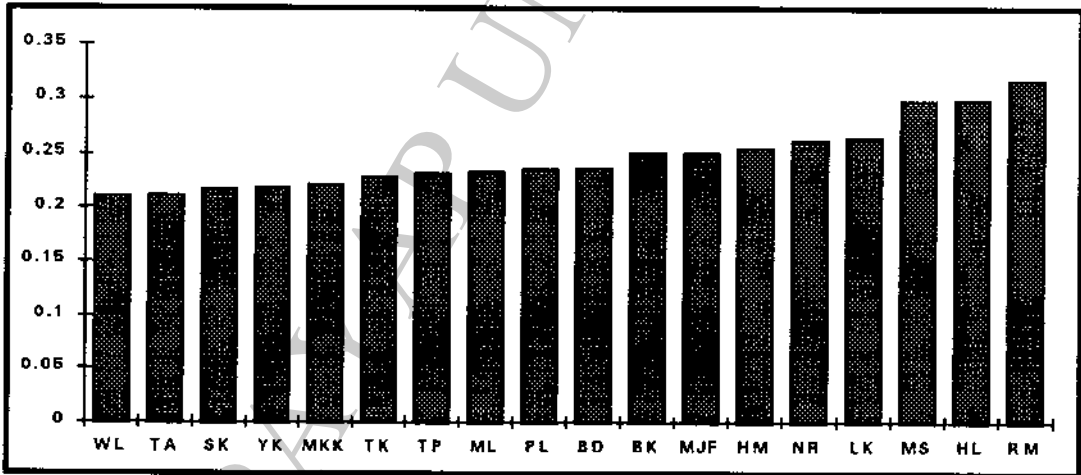
FIGURE 2 Phonological Relationships to the Entire Corpus

It is obvious from Figure 2 that significant skewing has occurred which favors the varieties that come from geographical areas that are better represented in the corpus.

Four of the five varieties that are "most typical" phonologically are from Chiang Mai: WANG LUANG, TIAN ANG, YANG KHROK, and MAE KHONG KLANG. To attempt to overcome this skewing I have also compared the phonology of each variety to the eight varieties that are most typical lexically. These eight represent two varieties from each of the four provinces where the most typical varieties are found.

Village	Ranking Against Corpus	Ranking Against "Top Eight"
SALOK	3	1
BAAN DOI	10	2
PA LAU	9	3
WANG LUANG	1	4
TIAN ANG	2	5
YANG KHROK	4	6
MAE LEK	8	7
MAE KHONG KLANG	5	8
THUNG KHONG	6	9
THUNG PHRAW	7	10
NONG RAK	14	11
LEKHO	15	12
MAE JONG FAI	11	13
BAAN KLANG	12	14
HUAI MUANG	13	15
HUAI LA	17	16
MAE SAAN	16	17
RUAM MIT	18	18

TABLE 9 A Summary of Phonological Similarity Rankings

There are a number of observations that can be made from Table 9. Comparing the two sets of rankings shown in this table, we notice that BAAN DOI is much more typical phonologically when it is compared to two typical varieties from each province than when it is compared to the entire corpus. In fact it becomes the second most typical after SALOK. The same can be said for PA LAU which moves from ninth to third in the ranking. WANG LUANG, on the other hand, appears less phonologically typical when the number of Chiang Mai varieties is limited to two. It has moved from first to fourth in the ranking.

Of the four varieties that were most similar in the lexicostatistics, all are in the

top ten phonologically using either system. SALOK and WANG LUANG are the two most "typical" varieties in their phonology with both of the them ranking very high whether compared to the entire corpus or to only the top eight (two from each province). The fact that SALOK, being located on the eastern border of the dialect grouping area is so typical phonologically is surprising. Unfortunately, I have not been able to obtain any migration information from the Phrae Pwo Karen which might clarify why this is so.

4.6 Identifying "Typical" Varieties of NORTHERN PWO KAREN

Although determining dialect centers simply on the basis of lexico- and phonostatistics is not possible, we can arrive at some conclusions concerning which varieties in this study are "typical" of NORTHERN PWO KAREN. In this section, we will look at the eighteen varieties in terms of both their phonological and lexical similarity to one another. If we combine the lexical rankings in Figure 1 with the phonological rankings in Table 9 (in which all members of the corpus are compared to the two most typical members from each of the four main provinces) the following pattern emerges.

Overall Ranking	Village	Lexical/Phono. Ranks	Average Ranks
1	BAAN DOI	1/2	1.5
2	WANG LUANG	2/4	3.0
3	SALOK	6/1	3.5
4	TIAN ANG	4/5	4.5
5	MAE LEK	3/7	5.0
6	PA LAU	7/3	5.0
7	YANK KHROK	10/6	8.0
8	MAE JONG FAI	5/12	8.5
9	THUNG KHONG	9/9	9.0
10	NONG RAK	8/11	9.5
11	THUNG PHRAW	13/10	11.5
12	BAAN KLANG	11/13	12.0
13	MAE KHONG K.	17/8	12.5
14	HUAI LA	12/15	13.5
15	HUAI MUANG	14/14	14.0
16	LEKHO	18/12	15.0
17	MAE SAAN	15/16	15.5
18	RUAM MIT	16/17	16.5

TABLE 10 A Summary of Phonological and Lexical Similarity Rankings

In Table 10 above, I have averaged the lexical and phonological rankings of each variety and given them an overall rank accordingly. Overall, BAAN DOI is the most typical of the varieties of NORTHERN PWO KAREN. The first ten varieties in Table 10 represent examples of NORTHERN PWO KAREN, with the higher ranked varieties being the most typical. NONG RAK is a less typical variety than was first thought, being tenth overall. YANG KHROK is seventh in the overall ranking substantiating it as a fairly typical variety of NORTHERN PWO KAREN. THUNG PHRAW, ranked eleventh overall, is a more atypical variety of NORTHERN PWO KAREN. BAAN KLANG, ranked number twelve, is the most closely related variety of LAMPANG PWO KAREN to NORTHERN PWO KAREN. Nevertheless, it is substantially different from the typical varieties of NORTHERN PWO KAREN. MAE KHONG KLANG is a step further removed from NORTHERN PWO KAREN than THUNG PHRAW. HUAI LA and HUAI MUANG can be viewed as members of another dialect grouping that have in many ways moved toward NORTHERN PWO KAREN through long-term contact. It is likely that further research will show that they have more similarities to a dialect grouping located south of the region where NORTHERN PWO KAREN is spoken.

LEKHO presents us with something of a quandary. Lexically it is closest to YANG KHROK at 93% lexical similarity, but overall it has only 89.3% lexical similarity with NORTHERN PWO KAREN. I will consider it to be a member of a different dialect grouping. It is likely that this grouping is most similar to the YANG KHROK variety of NORTHERN PWO KAREN.

4.7 Conclusions

We have seen considerable lexical and phonological evidence for positing two major dialect groupings of Pwo Karen in northern Thailand: NORTHERN PWO KAREN and LAMPANG PWO KAREN. In this final sub-section, I will attempt to chart the overall

relationships among the speech varieties covered in this study, from the point of view of both phonology and lexicon and will provide a map showing the areas where NORTHERN PWO KAREN and LAMPANG PWO KAREN are spoken.
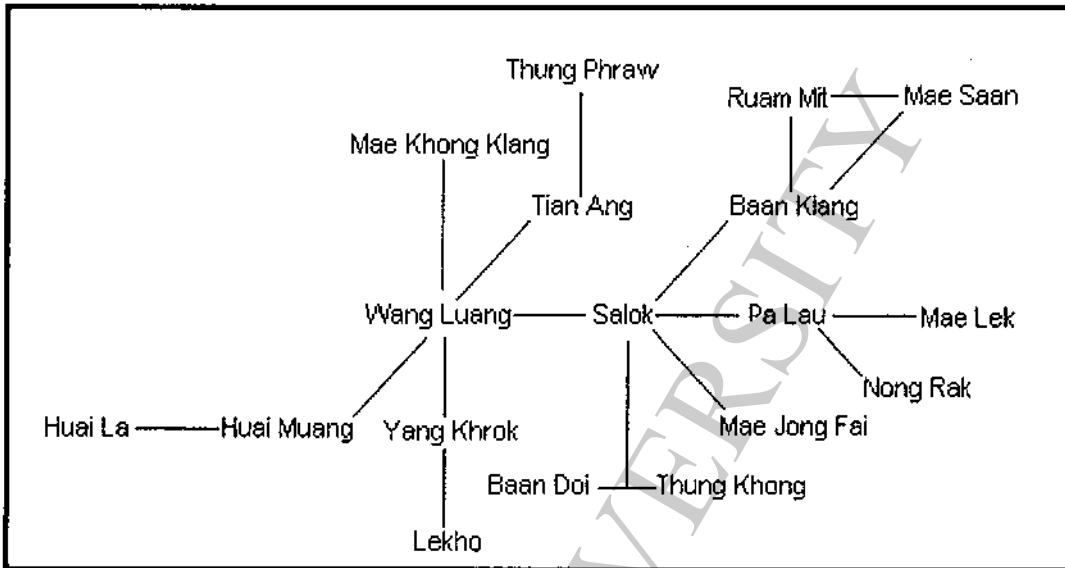
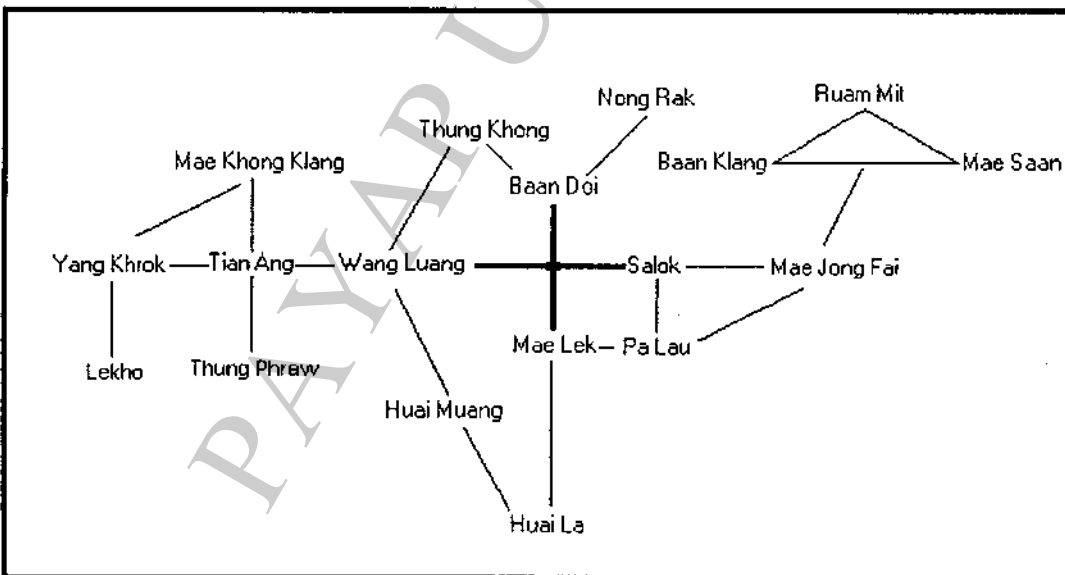**FIGURE 3   A Schema of Phonological Relationships**

FIGURE 4 A Schema of Lexical Relationships

Figures 3 and 4 illustrate the relationships among the various varieties of Pwo Karen considered in this study. Each variety is connected by a line to the other varieties to which it is most closely related. In Figure 4, the varieties that are the most typical lexically are shown branching from a hypothetical prototypical variety of NORTHERN PWO KAREN. In general terms, the more branches that separate a given variety from the center of the diagram the farther it is from the hypothetical dialect center. (In the case of the HUAI LA and HUAI MUANG varieties this is not clear in the diagram.)

Figure 4 illustrates an interesting point regarding the Chiang Mai and Mae Hong Son varieties. There appear to be several branches leading away from NORTHERN PWO KAREN at the TIAN ANG variety. The middle branch leads to YANG KHROK, which is a fairly typical variety of NORTHERN PWO KAREN. However, branching off from YANG KHROK is the LEKHO variety, which appears to be part of another dialect grouping. The upper branch connects YANG KHROK to MAE KHONG KLANG, which appears to be a very atypical variety of NORTHERN PWO KAREN. Finally, the lower branch connects YANG KHROK to THUNG PHRAW which also appears to be an atypical variety of NORTHERN PWO KAREN.

In Map 4 below, I have shown the locations of all of the word list locations once again and shaded areas in which NORTHERN PWO KAREN and LAMPANG PWO KAREN are spoken. The shaded areas are approximate and should not be viewed as rigid boundaries. Particularly in Chiang Mai province, we will find occasional varieties from other dialect groupings spoken within the proposed boundaries for NORTHERN PWO KAREN.
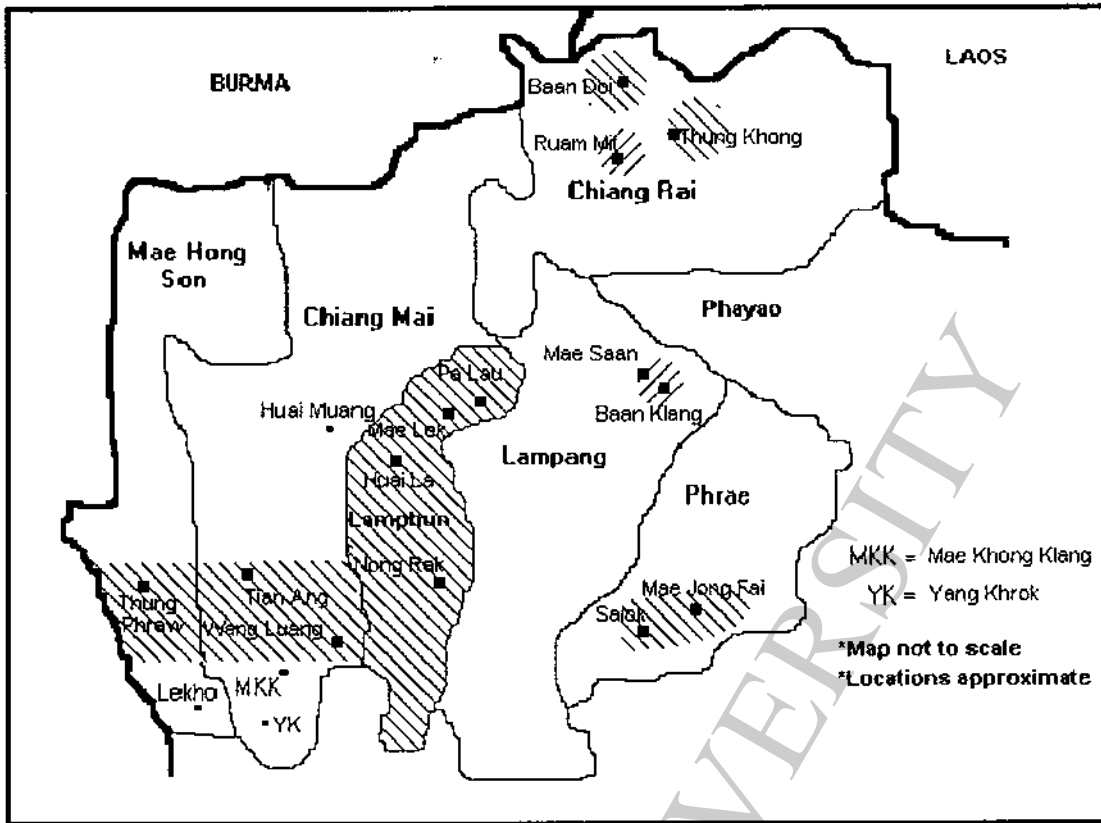
MAP 4   Areas where NORTHERN and LAMPANG PWO KAREN are Spoken

5. A Comparison of the Varieties of Pwo Karen Spoken in Northern Thailand to Varieties Spoken in Other Regions

In the previous sections I have made the claim that there are two major dialect groupings of Pwo Karen spoken in northern Thailand. One of these, I have labeled LAMPANG PWO KAREN, the other I have given a more general label: NORTHERN PWO KAREN. Labeling a dialect grouping "Northern" Pwo Karen implies a distinction between this dialect grouping and varieties spoken in other parts of Thailand. In this section, I will show that such a distinction is apparent from the data. The varieties of Pwo Karen spoken in the northern provinces are substantially different from those spoken in the central provinces of Kanchanaburi and Uthai Thani and the southern province of Phetburi. We will also see that it is quite different from three varieties spoken in Burma.

Data for the lexicostatistic analysis presented below were taken from a variety of

43

sources. The figures given are subject to a higher range of error than the figures presented for the varieties of Pwo Karen spoken in the North since I have limited knowledge of the Pwo Karen varieties spoken south of Mae Hong Son and elicitation techniques may have varied considerably between the various researchers. Of the comparisons below, those given for the dialects spoken in Phetburi and Kanchanaburi provinces have the lowest range of error since the data from Phetburi (PHU PHLU and NONG POEN TAEK) and Kanchanaburi (ONGSIT and WACHONGKHU) was gathered using the same word lists (elicited by David and Suree Anderson) which were collected quite recently (1987). In each case, 349-384 words were compared with the northern varieties. Although the data from Uthai Thani appear to be of the same variety as that from Kanchanaburi and Phetburi (data from Kaewsilpa, 1982), for comparison with the northern dialects only 232-237 words were available. The data from the KAWAREIK dialect in Burma come from a word list collected by Paulette Hopple in 1984. This variety is spoken in the mountain region of Burma between Moulmein and Mae Sot (Tak province of Thailand). For comparison between this variety and the northern varieties 222-225 words were available. Finally, the Pwo Karen data from BASSEIN and MOULMEIN were taken from Jones (1961). 222-227 words were compared with the northern varieties.