CHAPTER 6

SOCIOLINGUISTIC SURVEY ANALYSIS II – DETERMINING GROUPINGS OF KHUEN VARIETIES AND RELATIONSHIPS BETWEEN KHUEN AND OTHER SWT LANGUAGES

This chapter presents the analysis of the data from the sociolinguistic survey relating to Goals 3 and 4. Section 6.1 presents a summary of the phonological features of the seven Khuen varieties recorded on the survey. Section 6.2 describes the process and results of a lexicostatistical comparison of the wordlists from the Khuen varieties on the survey with other wordlists from SWT languages of the region. Section 6.3 discusses relationships between Khuen varieties. Section 6.4 draws conclusions relating to Goal 3 which seeks to determine the most suitable variety for use as a written standard. Sections 6.5 and 6.6 respectively present discussion of evidence and conclusions that can be drawn from the evidence gathered in this thesis concerning the relationships between Khuen and other regional SWT languages. Section 6.7 presents an evaluation of the instruments used on the survey.

6.1 Phonological Analysis of Wordlists

In this section a summary of the phonological features of the Khuen varieties sampled on the survey is presented. A fuller analysis including contrastive evidence for the existence of each phoneme for each variety is presented in Appendix 8. The general picture is one of great homogeneity across the seven varieties, although there is still some variation which will be discussed in the following sections, first for the consonants, then the vowels and finally the tones.

6.1.1 Consonants

The initial consonant phonemes of the seven Khuen varieties recorded on the survey are presented in Table 57.

Initial Consonant Phonemes		Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
Labial	pʰ	✓	✓	✓	✓	✓	✓	✓
	p	✓	✓	✓	✓	✓	✓	✓
	b	?	?	X?	✓	✓	✓?	X?
	f	X	X	✓?	✓	✓	X	X?
	m	✓	✓	✓	✓	✓	✓	✓
	w	✓	✓	✓	✓	✓	✓	✓
Alveolar	tʰ	✓	✓	✓	✓	✓	✓	✓
	t	✓	✓	✓	✓	✓	✓	✓
	d	X	✓?	X	✓?	✓?	X?	X?
	s	✓	✓	✓	✓	✓	✓	✓
	n	✓	✓	✓	✓	✓	✓	✓
	r	X	X	X	X	X	X	X
	l	✓	✓	✓	✓	✓	✓	✓
Post alveolar	c	✓	✓	✓	✓	✓	✓	✓
Palatal	j	✓	✓	✓	✓	✓	✓	✓
Velar	kʰ	✓	✓	✓	✓	✓	✓	✓
	k	✓	✓	✓	✓	✓	✓	✓
	ŋ	✓	✓	✓	✓	✓	✓	✓
Glottal	ʔ	✓	✓	✓	✓	✓	✓	✓
	h	✓	✓	✓	✓	✓	✓	✓
Total	✓	16	16	16	18	18	16	16
	✓?	0	1	1	1	1	1	0
	?	1	1	0	0	0	0	0
	X?	0	0	1	0	0	1	3
	X	3	2	2	1	1	2	1
Velar cluster	kw	X	X?	X?	✓?	✓	✓?	✓

Table 57 Initial consonant phonemes of seven Khuen varieties

For ease of comparison with previous Khuen research a row is included in Table 57 for every consonant phoneme identified by Egerød (1959:125). An extra row is appended for the only cluster observed in the Khuen varieties in the survey, namely the velar cluster /kw/. However, since Table 2 did not feature such a row,

the column totals at the bottom of Table 57 do not include the entries from the velar cluster row.

The evidence for most phonemes is clear-cut. In Table 57 the symbol ✓ represents those cases where there is clear evidence for the existence of a particular phoneme whereas X represents those cases where there is no evidence for a particular phoneme. Certain relatively low-frequency phonemes present mixed evidence as to their status in some speech varieties. In the light of previous Khuen research (Egerød 1959; Gedney [1964] 1994; Rasi 1978) it would appear that a process of change is underway with certain phonemes being gradually subsumed by other phonemes. Where the evidence is mixed, the symbol ✓? indicates that the evidence is inconclusive but over 50% of the evidence suggests that the phoneme exists. The symbol X? on the other hand signals inconclusive evidence but over 50% of the evidence suggests that the phoneme does not exist. Where 50% of the evidence suggests the phoneme exists and 50% suggests it does not exist, the symbol ? is used. For the sake of consistency with Table 6, the same phoneme labels are used. The rule in (3) summarises the conditioned variation for all varieties for the phoneme 'c'[35].

(3) Conditioned variation rule for 'c'

/tɕ/ → [ts] / _ low central vowel

[tɕ] / elsewhere

Note that the two varieties from Murng Lang village tract, namely Yang Kway and Wan Jay show similar patterns – as one might expect from their geographical proximity. There is strong positive evidence for 18 phonemes in these varieties whereas other varieties show such evidence for only 16 phonemes. These two varieties also show weaker positive evidence for one further phoneme yielding a

[35] This thesis follows Egerød (1959:125) and Rasi (1978:10) in using 'c' as a convenient single-character label for the phoneme whose phonetic realisation is either [tɕ] or [ts]. The reason for interpreting this as a single unit rather than a sequence of two consonants is that such an interpretation better fits the established syllable patterns.

total of 19 phonemes in agreement with Gedney ([1964] 1994). These two also show evidence of the velar cluster /kw/. There is no evidence of /r/ as a separate phoneme in any of these seven varieties. Perhaps the only other thing that may be inferred from the comparison of initial consonants in Table 57 is that Yang Lorh has the strongest evidence of loss of phonemes. This is in keeping with reported information that there has been language shift away from Khuen in the village. It should be noted however that Pa Jahm and Wan Jorhn also show strong evidence of loss of phonemes and no language shift was reported in those villages.

As far as final consonants are concerned, there is general agreement between the varieties, as can be seen by the entries in Table 58. As discussed in Appendix 8 there is meagre evidence for a glottal stop final, so all of the results in the 'Glottal' row are somewhat tentative.

Final Consonant Phonemes		Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
Labial	p	✓	✓	✓	✓	✓	✓	✓
	m	✓	✓	✓	✓	✓	✓	✓
	w	✓	✓	✓	✓	✓	✓	✓
Alveolar	t	✓	✓	✓	✓	✓	✓	✓
	n	✓	✓	✓	✓	✓	✓	✓
Palatal	j	✓	✓	✓	✓	✓	✓	✓
Velar	k	✓	✓	✓	✓	✓	✓	✓
	ŋ	✓	✓	✓	✓	✓	✓	✓
Glottal	ʔ	✓	✓	✓?	✓	✓	✓?	✓
Total		9	9	9	9	9	9	9

Table 58 Final consonant phonemes of seven Khuen varieties

6.1.2 Vowels

Appendix 8 presents evidence and a discussion of what conclusions can be drawn from the wordlist data concerning vowel phonemes in the seven Khuen varieties. A summary of the conclusions is given in Table 59. A row is included in the table for each vowel phoneme identified by Egerød (1959:125). The convention is adopted that if, for a given variety, only one phoneme exists at a particular

position in the vowel space, it will be labelled with the short form, even if the phonetic realisation is long. For example, /ɤ/ is always realised as [ɤː].

Vowel Phoneme	Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
i	✓	✓	✓	✓	✓	✓	✓
iː	X	X	X	X	X	X	X
ɯ	✓	✓	✓	✓	✓	✓	✓
ɯː	X	X	X	X	X	X	X
u	✓	✓	✓	✓	✓	✓	✓
uː	X	X	X	X	X	X	X
e	✓	✓	✓	✓	✓	✓	✓
eː	✓?	X	✓?	X	X	✓?	✓?
ɤ	✓	✓	✓	✓	✓	✓	✓
ɤː	X	X	X	X	X	X	X
o	✓	✓	✓	✓	✓	✓	✓
oː	✓?	X	✓?	X	X	✓?	✓?
ɛ	✓	✓	✓	✓	✓	✓	✓
ɛː	✓?	✓?	✓?	✓?	✓?	✓?	✓?
a	✓	✓	✓	✓	✓	✓	✓
aː	✓	✓	✓	✓	✓	✓	✓
ɔ	✓	✓	✓	✓	✓	✓	✓
ɔː	✓?	✓	✓?	✓?	✓?	✓?	✓?
Total ✓	10	11	10	10	10	10	10
Total ✓?	4	1	4	2	2	4	4
Total X	4	6	4	6	6	4	4

Table 59 Vowel phonemes of seven Khuen varieties

In Table 59 the ✓ symbol signifies that there is strong evidence for that particular phoneme in the data. An X on the other hand signifies that there is strong evidence against the existence of a particular phoneme. A question mark signifies that the evidence for that particular phoneme is not conclusive and so the proposed existence (✓) or non-existence (X) of a particular phoneme should be considered to be a hypothesis that needs to be confirmed or disproved by further data.

As described in Appendix 8, variation in phonetic length among the high vowels can be explained by conditioned variation rules, so length is not phonemic for the

high vowels. As far as the mid vowels are concerned, Table 59 shows that there are differences among the varieties for /e/ and /o/ whereas all varieties show a common pattern for /ɤ/, namely that there length is not phonemic because the phonetic realisation is always [ɤ:]. Insight into the differences between the varieties comes from considering the proto-vowels from which the modern reflexes derive. The varieties Yang Lorh, Wan Jorhn, Wan Kahng and Murng Jem display the same modern reflex, namely /e/, of the PSWT vowel *e. The varieties Pa Jahm, Yang Kway and Wan Jay on the other hand have the reflex [ɛ]. Moreover Pa Jahm, Yang Kway and Wan Jay have no short form of the close-mid front unrounded vowel [e:]. A similar situation pertains for modern reflexes of PSWT *o, i.e., Pa Jahm, Yang Kway and Wan Jay have the reflex [ɔ] whereas Yang Lorh, Wan Jorhn, Wan Kahng and Murng Jem have the reflex [o]. Pa Jahm, Yang Kway and Wan Jay have no short form of the close-mid back rounded vowel [o:].

As described in Appendix 2 the Khuen 406 wordlist included items to check whether Khuen has the 'nasal umlaut' feature as found in Lue. When preceding a nasal, contemporary Lue varieties have [u] and [i] as reflexes of PSWT *o and *e. There is no significant evidence of a nasal umlaut in Khuen.

6.1.3 Tones

A summary of the distribution of the five tonemes is presented in Table 60.

Tone Box Ref	Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
A1	1	1	1	1	1	1	1
A2							
A3	2	2	2	2	2	2	2
A4							
B1	3	3	3	3	3	3	3
B2							
B3							
B4							
C1	4	4	4	4	4	4	4
C2							
C3							
C4	5	5	5	5	5	5	5
DS1	2	2	2	2	2	2	2
DS2							
DS3							
DS4	5	5	5	5	5	5	5
DL1	3	3	3	3	3	3	3
DL2							
DL3							
DL4							

Table 60 Tone distribution of seven Khuen varieties

The bold lines in the table indicate the tone splits and it is clear from the table that there is almost universal agreement across the seven varieties. One curious exception is the split in the A column for Murng Jem – signified by the dotted line in the table. The evidence is inconclusive as to where to place the tone split with some items in box A3 having toneme 1 and other items having toneme 2. Since there are no obvious phonological factors that could be conditioning this variation it might be due to language contact with Shan which has the tone split between boxes A3 and A4. All the other Khuen varieties clearly have the split between A2 and A3. It must be stressed that this is merely a hypothesis and further research is necessary to investigate whether this pattern is true for the whole community and if so what factors can best account for it. It should be noted however, that the Shan interpreter on the survey fieldwork team observed that Murng Jem subjects spoke like Shan speakers, as did subjects in Yang Lorh.

In the review of previous Khuen phonology research in Chapter 2 the main difference between the tone systems was whether or not there is a tone split in the B column: some studies showed a B123-4 pattern while others showed a B1234 pattern. As can be seen from Table 60 for all seven varieties there is no evidence of a tone split in the B column. Moreover, this is also the case for the DL column and the phonetic similarity of the tone found in column DL to that found in column B leads to the conclusion that B=DL for all seven varieties.

6.2 Lexicostatistical Comparison of Wordlists

6.2.1 Consolidation of data

Having analysed the phonological features of each variety individually, it is possible to compare different varieties. In light of the discussion in Section 3.1.2 about comparing lexical items with the same contemporary meaning, great care was taken in choosing which lexical items to include in the comparison. The primary limitation on the items available for inclusion arose from the fact that the data were from several different sources and came in several different formats. These sources and formats are described in Table 61.

Ref	Variety Name	Location	Data source	Format	Year
1	Standard Thai	Bangkok	Haas (1964)	Dictionary	2007
2	Northern Thai	Chiang Mai	Author - Primary	Khuen 406 wordlist	2007
3	Khuen	Yang Lorh	Author - Secondary	Khuen 406 wordlist	2006
4	Khuen	Pa Jahm	Author - Secondary	Khuen 406 wordlist	2006
5	Khuen	Wan Jorhn	Author - Secondary	Khuen 406 wordlist	2006
6	Khuen	Yang Kway	Author - Secondary	Khuen 406 wordlist	2006
7	Khuen	Wan Jay	Author - Secondary	Khuen 406 wordlist	2006
8	Khuen	Wan Kahng	Author - Secondary	Khuen 406 wordlist	2006
9	Khuen	Murng Jem	Author - Secondary	Khuen 406 wordlist	2006
10	Khuen	Wan Bo	Author - Primary	MSEAG 436 wordlist	2004
11	Khuen	Lah Murng	Author - Primary	MSEAG 436 wordlist	2003
12	Lue	Murng Yorhng	Hudak (1994)	Glossary	1964
13	Lue	Jinghong	Hudak (1994)	Glossary	1964
14	Shan	Keng Tung	Author - Primary	Khuen 406 wordlist	2007
15	Mao	Nam Kham	Apiradee (2007)	EAG 550 wordlist	2004

Table 61 Sources of data for lexicostatistics

The third column of Table 61 lists the hometown of the LRP(s) who provided the data. The entries in column 4 show that most of the data is from fieldwork in

which the author has been involved, either in the secondary sense of initiating the data collection, such as with the present survey, or in the primary sense of eliciting and recording the data. In each case the data was recorded and the sound files analysed by the author. The Mao data was collected in 2004 and analysed by Apiradee (2007). The sound files were reanalysed by the author for this study to ensure consistency of transcription with other varieties. The Lue data was collected by Gedney in 1964 but it was not published until 1994 with Hudak as editor. The final column gives either the year in which the data was collected or in which the author verified with a native speaker that it still represents contemporary usage.

The first step in the comparison was to determine the subset of words that were common to all the lists – some 184 items. Each of these common items was then examined for each variety to rule out any inconsistencies. The final list comprised 157 items. The author believes that the effort put in to rule out inconsistencies, which incorporated as appropriate the wealth of information on SWT languages encapsulated in the historical reconstruction of the proto forms, gives the lexical similarity percentages calculated from the data greater reliability than would otherwise be the case (Simons 1977b:81). There were two types of inconsistency that caused items to be excluded from the final list.

Firstly an item was excluded if the given word in any of the varieties was suspect. For example the entry for item 59 'bamboo' from Pa Jahm village was [maːj^{51}] which is the generic word for 'wood'. This item was therefore removed from every list. By removing such items from every list a lot of perfectly good data is being disregarded. Notwithstanding this approach was chosen in the interests of simplicity and consistency – the lexical similarity percentage of any two varieties is thus based on exactly the same set of lexical items. It could be argued that this is 'interfering' with the data, but such 'pre-processing' is in line with Rensch (1992:14) who found that lexical similarity percentages were typically lowered by

between 5 and 10 percentage points if such 'pre-processing' of the items on the list was not carried out.

The second reason for which items were excluded from the list was probable ambiguity. A clear-cut example was item 270 'return'. The Khuen 406 wordlist further specified this item to refer to somebody returning to their home. The EAG 550 wordlist however was not so specific and it appears that the word meaning 'to return something borrowed' was elicited. Another example was item 136 'frog'. There are many types of frog: some are poisonous whereas others are regularly eaten in some of the cultures included in this study. Moreover, there are different words for different types of frogs and so unless the elicitation process specifically leads the LRP to a particular choice, the LRP will choose which word to give. It is therefore quite likely that different LRPs will give different words – especially if different researchers are conducting the fieldwork. Items were only excluded if there was good evidence to suggest the possible contemporary use of more than two words for a particular concept. In this respect the reconstructed PT forms provide a great reference since they represent the synthesis of lexical and phonological information from a very extensive range of Tai languages. In the case of item 136 'frog' Li (1977:324) had two reconstructed forms for the gloss 'frog', namely *kop (tone D1S) and *khiat (tone D1L).

The effect of excluding these 'ambiguous' items from the list is likely to raise the lexical similarity percentages. The author contends that this is not an artificial raising but an attempt to ensure that the percentages are not unnecessarily lowered in the sense described by Rensch (1992:14). Moreover even if the application of this principle of excluding 'ambiguous' items introduces a measure of bias, such upward bias would tend to affect all varieties indiscriminately. Recall that the purpose for which lexicostatistics is being used in this work is to investigate groupings both within the Khuen language as well as in the wider context of SWT. The benefit of including languages that are expected to be more distantly related to Khuen such as Standard Thai and Tai Mao is to give a sense of perspective to the relative closeness of the Khuen varieties themselves. Although

not specifically at issue in this work, the level of intelligibility between the different speech varieties is of interest. Nahhas (2007b:81) points out that when using lexical similarity as a screen for a lack of intelligibility, if there are doubts about whether or not to count two lexical items as similar or not, the researcher should classify them as similar – in other words an upward bias – to avoid wrongly concluding that two varieties are unintelligible.

6.2.2 Calculating the lexicostatistic similarity percentages

The Blair-type criteria used to determine whether or not two words are lexically similar are given in Appendix 9. Two lexical similarity matrices were computed, one using all 157 items common to all data sets and the other using only the 100 highest ranked[36] items among them. The lexical similarity percentages based on the 100 items are presented in Figure 44.

1	100													
2	91	100												
3	81	90	100											
4	81	93	95	100										
5	***80***	92	96	97	100									
6	***80***	92	95	98	99	100								
7	81	92	96	98	99	**100**	100							
8	***80***	92	95	98	99	**100**	**100**	100						
9	81	93	94	97	98	99	99	99	100					
10	82	94	94	96	97	98	98	98	99	100				
11	83	95	93	95	96	97	96	97	98	99	100			
12	86	95	92	94	93	94	94	94	94	93	95	100		
13	86	95	92	94	93	94	94	94	95	94	95	**100**	100	
14	81	90	97	93	94	93	94	94	94	94	93	92	93	100
15	***80***	87	93	89	90	89	90	90	90	90	89	88	88	95	100
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

Figure 44 Lexical similarity percentages for 15 Tai varieties (top 100 items)[37]

[36] The ranking was that developed by Mann (2004) as used in the design of the Khuen 406 wordlist.

[37] The numbering key is as follows: 1-Thai; 2-Northern Thai; 3-Yang Lorh; 4-Pa Jahm; 5-Wan Johrn; 6-Yang Kway; 7-Wan Jay; 8-Wan Kahng; 9-Murng Jem; 10- Wan Bo; 11- Lah Murng; 12-Lue Murng Yorhng; 13-Lue Jinghong; 14 Shan Keng Tung; 15-Mao Nam Kham

The varieties are labelled by the same reference numbers as in Table 61. The most striking thing about the percentages is how high they are. The similarity percentages for the 9 Khuen varieties (seven from the survey plus two previous wordlists by the author) are shaded grey in the matrix. These percentages include three 100% entries which are highlighted in boldface. One further 100% entry is found for the similarity percentage between the two Lue varieties (numbers 12 and 13 in the matrix.) The lowest numbers in the matrix – shown in boldface italic – are four entries of 80% in the first column. In fact the lowest value in each row lies in the first column signifying that Standard Thai generally has the weakest relationships with the other varieties.

To assist with the interpretation of the lexical similarity percentages in Figure 44, Table 62 shows how large a difference is needed to be significant at the 15% and the 5% levels respectively. The significance levels assume that the wordlist data has high reliability, which the author asserts to be the case here.

Similarity percentage	Significant at 15% level	Significant at 5% level	Similarity percentage	Significant at 15% level	Significant at 5% level
80	83	85	90	93	94
81	84	86	91	94	95
82	85	87	92	94	95
83	86	88	93	96	96
84	87	89	94	97	97
85	88	90	95	97	98
86	89	90	96	98	99
87	90	91	97	99	99
88	91	92	98	99	100
89	92	93	99	100	100

Table 62 Significant differences for 100 item wordlists (adapted from Simons 1977b:97)

Figure 45 shows the phenogram of the structure produced by the UPGMA algorithm[38] from the data in Figure 44. Note that the length of horizontal line from the name of the speech variety to the vertical line representing a join with

[38] The Unweighted Pair-Group Method using Arithmetic Average algorithm is a type of cluster analysis.

another speech variety is related to the degree of similarity between those two varieties. For example the two Lue varieties have a similarity of 100% so they are joined by a vertical line – the notional horizontal line is of length zero. Contrast this with the length of the horizontal line joining Thai to the other varieties – the line is the longest signifying that Thai has the lowest similarity percentages with the other varieties.

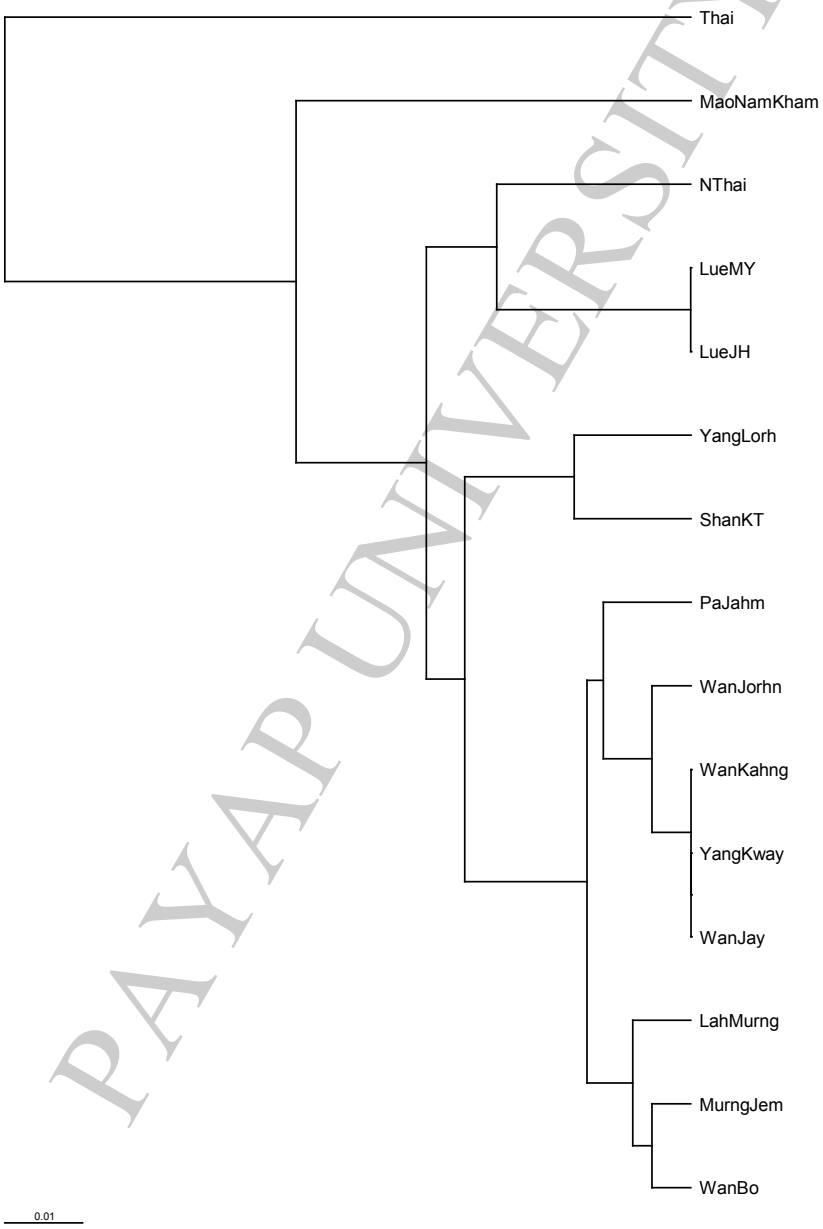

Figure 45 Phenogram showing relationships between 15 Tai varieties (100 wordlist items)

Examining the subgroupings in Figure 45 it is seen that the Khuen varieties are generally together in the bottom two thirds of the diagram. The horizontal lines joining the subgroups of varieties are generally short indicating that the lexical similarity percentages are generally high between Khuen varieties. The values in Table 62 are used to evaluate the relationships depicted in Figure 45. For example the lexical similarity percentages between Yang Lorh and Pa Jahm (95%) and Shan and Pa Jahm (93%) are not significantly different. On this basis Yang Lorh and Shan are correctly placed in the same sub-group. Further, the similarity percentage between Pa Jahm and Murng Jem (97%) and that between Yang Lorh and Murng Jem (94%) can be seen to be significant at the 5% level. On this basis Yang Lorh and Pa Jahm are correctly placed in different sub-groups. It would be cumbersome to discuss the statistical significance of every possible sub-group in Figure 44. Suffice to say that as the foregoing example shows, even the relatively fine sub-grouping of Khuen varieties can be shown to have some statistical support.

The lexical similarity percentages based on all 157 items are shown in Figure 46.

1	100														
2	89	100													
3	80	89	100												
4	77	89	94	100											
5	78	90	96	96	100										
6	78	90	96	97	98	100									
7	78	89	96	96	97	99	100								
8	78	89	96	97	98	99	99	100							
9	78	90	94	96	98	97	97	97	100						
10	79	90	95	95	96	98	97	97	97	100					
11	80	92	94	94	96	97	96	96	96	99	100				
12	84	94	94	93	93	95	94	94	93	94	96	100			
13	84	93	94	93	93	95	94	94	94	95	96	**100**	100		
14	77	85	94	90	91	92	93	92	90	92	90	90	90	100	
15	***75***	82	89	87	87	87	89	87	87	87	86	85	85	91	100
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**

Figure 46 Lexical similarity percentages for 15 Tai varieties (all 157 items)

The percentages are generally lower than in Figure 44 but are still generally high. The Khuen varieties are shaded in grey. The highest similarity percentage is 100% between the two Lue varieties (12 and 13). This is shown in boldface in the table. The lowest number is 75% between Standard Thai and Mao Nam Kham.

To assist with the interpretation of the lexical similarity percentages in Figure 46, Table 63 shows how large a difference is needed to be significant at the 15% and the 5% levels respectively. The significance levels assume that the wordlist data has high reliability, which the author claims to be the case here. Note that the values in the tables are for 150-item wordlists and our lexical similarity percentages are based on 157-item lists. The effect of this is small and will not be discussed further.

Similarity percentage	Significant at 15% level	Significant at 5% level	Similarity percentage	Significant at 15% level	Significant at 5% level
75	78	80	90	92	93
76	79	80	91	93	94
77	80	81	92	94	95
78	81	82	93	95	96
79	82	83	94	96	97
80	83	84	95	97	97
81	84	85	96	98	98
82	85	86	97	98	99
83	86	87	98	99	100
84	87	88	99	100	100
85	88	89			
86	89	90			
87	89	91			
88	90	91			
89	91	92			

Table 63 Significant differences for 150 item wordlists (adapted from Simons 1977b:98-99)

Figure 47 shows the phenogram of the structure produced by the UPGMA algorithm from the data in Figure 46.

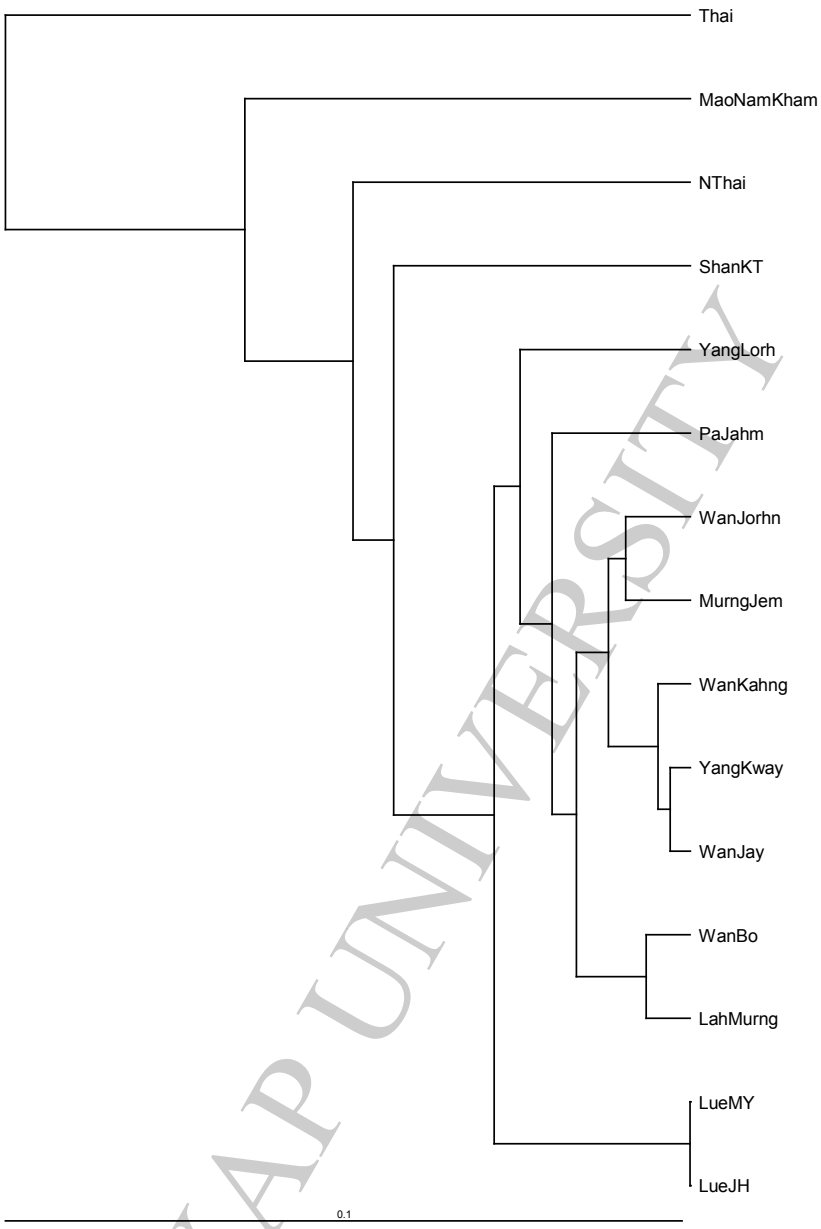

Figure 47 Phenogram showing relationships between 15 Tai varieties (157 wordlist items)

Examining the subgroupings in Figure 47 it can be seen that the Khuen varieties are generally together in the bottom two thirds of the diagram. Moreover, since the lexical similarity percentages are based on a larger number of lexical items, the size of the difference required to be statistically significant is generally smaller

than for the 100-item wordlist discussed above so even the fine sub-grouping of Khuen varieties has some statistical support.

It is worth noting that demonstrating that a particular feature – in this case the difference between two lexical similarity percentages – has statistical significance is not the end of the discussion. One must also state whether that feature is of practical importance. As discussed in Section 3.1.2 there is no agreement on the practical importance of small differences in lexical similarity percentages. Conclusions about the fine sub-grouping of Khuen varieties cannot be drawn on the basis of lexical similarity percentages alone. Rather evidence is sought from comparison of phonological features as well as the perceptions of native speakers to corroborate the tentative findings based on lexical similarity.

The following two sections discuss the relationships between Khuen varieties themselves and the relationships with other neighbouring Tai varieties.

6.3 Relationships between Khuen Varieties

In promoting Khuen literacy it is important that the variety being promoted is both *acceptable* and *accessible* to the people it is intended to benefit. Acceptability is gauged in terms of the attitudes of Khuen speakers in general to the chosen variety. Accessibility relates to the phonological, lexical and grammatical closeness of the chosen variety to all of the other varieties. In attempting to address these issues, the following three sections present evidence of groupings of Khuen varieties based on lexical similarity, phonological criteria and speaker perceptions. These analyses give a sense of the overall variability of Khuen varieties which provides a useful perspective on the selection of any single variety for use as a written standard. Section 6.3.5 investigates whether Khuen speakers generally acknowledge one variety as pre-eminent. If this is the case, then that particular variety is obviously a leading candidate for use as a written standard, provided that it is also reasonably close to other varieties phonologically, lexically and grammatically.

6.3.1 Groupings based on lexical similarity

This section seeks to answer the following research question:

Research Question 3.1: What are the groupings of Khuen varieties based on lexical similarity?

Perhaps the main thing to be stated is that because of the generally high level of the lexical similarity percentages, care must be taken to check whether or not the difference between two percentages is significant. This leads on to the question of whether a difference between two similarity percentages that are statistically significant is of practical importance. Since interpreting small differences between lexical similarity percentages is fraught with difficulties for reasons laid out in Chapter 3, the groupings derived by lexicostatistical comparison are simply compared with those derived by comparison of phonological features. The groupings are further compared with the distribution of other factors that might affect the similarity of two varieties such as geographical proximity or reported language contact.

The groupings of Khuen varieties based on lexical similarity percentages among the top 100 items were pictured in Figure 45. The Khuen varieties are generally grouped together representing the fact that the lexical similarity percentages are all 93% or higher. Yang Lorh is grouped with Shan, which might be expected given the apparent shift to Shan as the main language used in Yang Lorh. Yang Kway is grouped together with Wan Jay, which is to be expected given the fact that both villages lie in Murng Lang village tract. Wan Kahng, which is in a village tract adjacent to Murng Lang, is also grouped with Wan Jay and Yang Kway.

The groupings of Khuen varieties based on lexical similarity percentages among the 157 items common to all lists were pictured in Figure 47. The Khuen varieties are generally grouped together representing the fact that the lexical similarity

percentages are all 94% or higher. Yang Kway is again grouped with Wan Jay and Wan Kahng.

Summarising the evidence in this section, there are two main points. Firstly, the generally high lexical similarity percentages among all Khuen varieties indicate that Khuen as a language enjoys a high degree of homogeneity. This is perhaps not surprising given the relatively small geographical area represented by the varieties in the sample. Moreover the lexical similarity percentages do not provide strong evidence of exact subgroupings of Khuen varieties. Secondly, Yang Lorh is marginal in relation to other Khuen varieties and is therefore not an obvious candidate for use as the standard variety.

6.3.2 Groupings based on phonological features

This section seeks to answer the following research question:

Research Question 3.2: What are the groupings of Khuen varieties based on phonological criteria?

When comparing the phonological features of Khuen varieties there is a great deal of homogeneity but some differences do exist which point to certain sub-groupings. Table 64 is a reduced version of Table 57 after all of the features which showed no variation across the varieties have been removed.

Initial Consonants		Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
Single phonemes	b	?	?	X?	✓	✓	✓?	X?
	f	X	X	✓?	✓	✓	X	X?
	d	X	✓?	X	✓?	✓?	X?	X?
Cluster	kw	X	X?	X?	✓?	✓	✓?	✓

Table 64 Initial consonant differences between Khuen varieties

The large number of question marks in the table indicates that the evidence is often not clear-cut. There is clear evidence however that Yang Kway and Wan Jay varieties can be identified as a sub-group. These varieties display the strongest

evidence among any of the varieties of retaining the phonemes /b/, /f/ and /d/ as well as the velar cluster /kw/. Together these features build up a picture of these two varieties as being more conservative than the other varieties. Moreover they both lie in Murng Lang village tract so one might expect them to group together. Pa Jahm shares one of these features with Yang Kway and Wan Jay, namely the voiced stop /d/ whereas Wan Kahng shows some evidence for the voiced stop /b/ and the velar cluster /kw/.

Table 65 is a reduced version of Table 59 after all of the features which showed no variation across the varieties have been removed. The entries in Table 59 show that Yang Kway, Wan Jay and Pa Jahm follow the same pattern. In this table all other varieties have an identical pattern but it should be remembered that this pattern is basically one of insufficient evidence to either confirm or rule out the existence of phonemic length. It is nevertheless clear that these varieties differ from Pa Jahm, Yang Kway and Wan Jay with respect to these vowel features.

Phoneme Label	Yang Lorh	Pa Jahm	Wan Jorhn	Yang Kway	Wan Jay	Wan Kahng	Murng Jem
e:	✓?	X	✓?	X	X	✓?	✓?
o:	✓?	X	✓?	X	X	✓?	✓?
ɔ:	✓?	✓	✓?	✓?	✓?	✓?	✓?

Table 65 Vowel differences between Khuen varieties

As mentioned in Section 6.1.2 modern reflexes of PSWT *e and *o in Yang Lorh, Wan Jorhn, Wan Kahng and Murng Jem varieties are [e] and [o] respectively whereas in Pa Jahm, Yang Kway and Wan Jay varieties the modern reflexes are [ɛ] and [ɔ] respectively.

Summarising the evidence from phonological features, Yang Kway and Wan Jay show similar patterns in consonants and vowels. This supports Rasi's (1978:2) assertion that a distinct Murng Lang variety exists, although certain of these features are also shared by Pa Jahm and Wan Kahng. There is not enough clear-cut evidence to say whether the varieties other than Yang Kway and Wan Jay

should all be grouped together or whether further subgroupings exist. It is clear however that Yang Lorh is not the best representative of these other varieties and hence not the ideal candidate for a standard variety. From their geographical location and these varieties are associated with the Kang Murng variety studied by Gedney ([1964] 1996) and Rasi (1978).

6.3.3 Groupings based on speaker perceptions

This section seeks to answer the following question:

Research Question 3.3: What groupings of Khuen varieties are perceived by Khuen speakers?

Subjects were asked to group villages according to whether people from those villages spoke Khuen the same as them, slightly differently or very differently. The analysis of the responses to these questions does not yield great insight. The villages specified tend to be very close to the village of the particular subject giving the answer. While there are many villages in common among the lists of subjects from the same village, there is almost no overlap between lists from subjects from different villages. This does not mean that there are 6 distinct varieties; rather it should be interpreted as a reflection of the fact that subjects' knowledge of other villages and accompanying linguistic variation is very limited to the area surrounding their home village. Of those subjects who specified a village not in the immediate vicinity of their own village, the Khuen varieties spoken in Murng Lang were mentioned by one subject from Yang Lorh; two subjects from Pa Jahm; and one subject from Wan Kahng as being 'very different' to their own variety of Khuen.

6.3.4 Conclusions about groupings of Khuen varieties

This section draws together the findings about groupings of Khuen varieties from the previous sections. The strongest evidence is for Yang Kway and Wan Jay to be grouped together. This is based on phonological evidence, lexical similarity

percentages and speaker perceptions. Moreover it correlates with geographical location and provides some support for Rasi's assertion that Murng Lang is a recognised dialect of Khuen (Rasi 1978:2).

Yang Lorh village is most closely related to Shan in terms of lexical similarity and shows most evidence of loss of phonemes compared to other Khuen varieties. This is consistent with the reports of shift to Shan among Yang Lorh residents. There is insufficient evidence to further sub-group the remaining varieties, namely Pa Jahm, Wan Kahng, Wan Jorhn and Murng Jem. Because of their geographical location they are associated with the Kang Murng variety studied by Gedney ([1964] 1994) and Rasi (1978).

6.3.5 Most prestigious variety

In attempting to weigh the relative merits of different varieties as candidates for use as a written standard, it is important to gauge speakers' perceptions of and attitudes towards the different varieties as well as examining their phonological features. This section seeks to answer the following research question:

> **Research Question 3.4:** Which Khuen variety has most prestige in the eyes of Khuen speakers?

Since this research question is attempting to gauge people's attitudes, subjects were asked related questions which reveal something of their attitudes towards and perceptions of Khuen varieties. Q47 attempts to probe whether Khuen speakers perceive one location to be home to the 'best'[39] variety of Khuen.

Q47 In which village is Khuen spoken best?

The responses to Q47 are presented in Table 66.

[39] This term is deliberately vague. When giving their answer, subjects were free to interpret the question as they pleased.

Subject's own village	Different village in same village tract	Yang Lorh	Wan Kahng	Murng Jem	Wan Loi	Murng Lang	Total
38 (69%)	5 (9%)	2 (4%)	2 (4%)	3 (5%)	2 (4%)	3 (5%)	55 (100%)

Table 66 Q47 In which village is Khuen spoken best?

The most striking thing to note from Table 66 is that 38/55 [69%] of the subjects who responded (not counting subjects from Yang Lorh) said their own village spoke Khuen the best. A further 5/55 [9%] named another village in the same village tract as their own. The final two data columns could be combined since Wan Loi is a village in Murng Lang village tract. One subject from Yang Kway village (also in Murng Lang village tract) commented that Yang Kway was the best variety because it was 'pure Khuen, spoken as the written form'. This perception of Yang Kway being closer to the written form is consistent with the findings in Section 6.1 regarding phonological features. For example in the Khuen orthography the initial consonant ဝ [ba?^{5}] is used for words reconstructed with PT *ʔb whereas the consonant ဝ [wa?^{53}] is used for words reconstructed with PT *w.

As can be seen from the comparison of initial consonants in Table 57, Yang Kway apparently preserves a phonemic distinction between /b/ and /w/ that matches this orthographic distinction. In fact the other village from Murng Lang village tract included in the survey, viz. Wan Jay, exhibits the same phonological features as Yang Kway. It is therefore likely that these phonological features are widespread in Murng Lang village tract. Moreover it is not surprising for Murng Lang varieties to be perceived as distinctive by subjects from other village tracts.

The responses to Q47 suggest that Khuen speakers' knowledge of other varieties is generally limited to the villages closest to their own. The numbers of subjects naming villages other than their own village as the place where Khuen is spoken best are too small to draw hard and fast conclusions. There is however a hint that

varieties in Murng Lang village tract are respected because of their closeness to the written form of Khuen.

The following question seeks to gain insight into attitudes towards the prestige of Khuen villages.

Q48 Which Khuen village is the most important?

The responses to Q48 are laid out in Table 67 and follow a similar pattern to those of Q47.

Subject's own village	Different village in same village tract	Yang Lorh	Kat Taw	Kat Fah	Total
43 (81%)	2 (4%)	4 (8%)	3 (6%)	1 (2%)	53 (100%)

Table 67 Q48 Which Khuen village is the most important?

Summing the numbers in the first two columns of Table 67 shows that, of the subjects who gave an answer, 45/53 [85%] named their own village or a village in the same village tract as their own. If this is taken at face value then the most obvious conclusion to draw is that Khuen speakers generally do not perceive one village as pre-eminent for the whole Khuen region. Yang Lorh village was named by 4/53 [8%] of subjects from villages other than Yang Lorh citing as reasons its proximity to Keng Tung; the quality of houses; the number of educated people; and the fact that a Thai princess used to visit the village. Kat Taw was named by 3/53 [6%] of the respondents citing as reasons its size, leadership and development.

While Q47 and Q48 were intended to allow subjects to name towns such as Keng Tung as for an answer, the Shan translation effectively precluded this, asking only for the most important village or the village where Khuen is spoken best. These questions therefore do not provide clear information about subjects' attitudes to Keng Tung.

To summarise the evidence in this section, no variety is shown to be the most prestigious. It is not clear whether this is due to lack of clarity in the questions, subjects' limited knowledge of other Khuen varieties or that there is in fact no single most prestigious variety. It may be that each of these factors makes a contribution.

6.4 Conclusions Relating to Goal 3

The study does not clearly identify any single variety as having the most prestige in the eyes of Khuen speakers. This means that determining the most suitable variety for use as a written standard must be based on the findings relating to groupings of Khuen varieties. Two groupings of varieties are identified, namely Murng Lang and Kang Murng. However, to put this choice in perspective, the generally high degree of homogeneity among varieties both phonologically and lexically means that any variety apart from Yang Lorh could reasonably serve as the standard. The greater geographical range of varieties associated with the Kang Murng variety grouping gives it an advantage over the more geographically limited Murng Lang variety grouping. However more information is needed particularly regarding the attitudes of Khuen speakers towards the different varieties before a conclusion can be drawn with any confidence.

6.5 The Place of Khuen within SWT

This section uses both phonological features and lexical similarity percentages to assess the position of Khuen within the group of SWT languages.

6.5.1 Groupings based on lexical similarity

This section seeks to answer the following question:

Research Question 4.1: Which Tai languages are most closely related to Khuen based on lexical similarity?

Both Figure 45 and Figure 47 show Standard Thai to be the language least closely related to the Khuen varieties. Mao Nam Kham and then Northern Thai are the next most distant languages. Whereas Figure 45 places Shan closest to the Khuen varieties, Figure 47 places the two Lue varieties closest to the Khuen varieties. Given the difficulties with interpreting the practical importance of differences between lexical similarity percentages described in Chapter 3, the groupings suggested by Figure 45 and Figure 47 should not be taken as conclusive. Rather corroborating evidence should be sought from comparison of phonological features, which is the subject of the following section.

6.5.2 Groupings based on phonological features

This section seeks to answer the following question:

Research Question 4.2: Which Tai languages are most closely related to Khuen based on phonological features?

Chamberlain's classification of SWT languages was based on a hierarchy of phonological features (Chamberlain 1975:50). Primary among these features was the so-called P-PH distinction referring to whether or not PT voiced stops developed into homorganic unaspirated or aspirated voiceless stops in modern varieties. Of the speech varieties considered in this study, Standard Thai alone falls into the PH Group and so would be separated off from the other varieties by this criterion. Robinson's subsequent work on Southwestern P-Group languages follows Chamberlain's general approach and identifies five criteria which consistently separate Tai Mao, Tai Nuea and Tai Khamti from other Southwestern P-Group languages (Robinson 1994:138). Among these criteria is the pattern of tone splits and mergers in the PT A column, which is also used to subdivide those varieties labelled 'Other' in Robinson's classification which is reproduced in Figure 48.

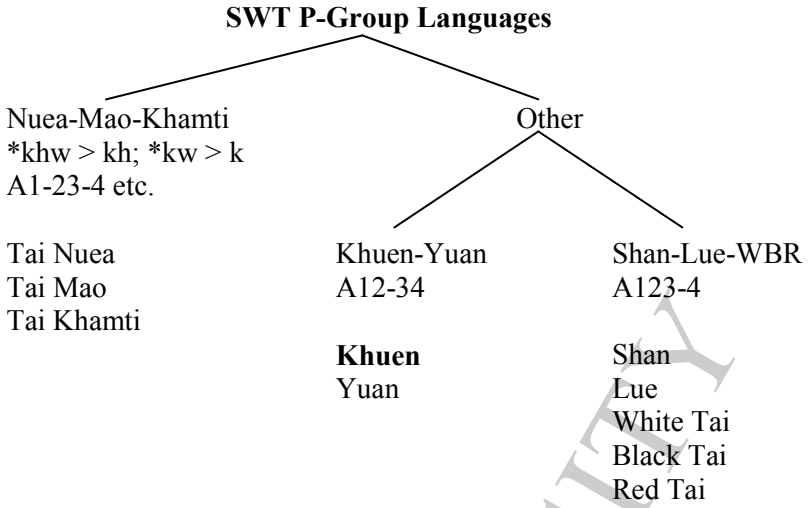

Figure 48 Classification of SWT P-Group languages (adapted from Robinson 1994:141)

As can be seen from Figure 48 the A-column tone split criterion separates Khuen and Northern Thai from Lue and Shan. In the present work however, there are two Lue varieties which would be classified into different subgroups by virtue of their different pattern of tone splits in the A column. To be precise, Lue Murng Yorhng would be grouped with Khuen and Northern Thai because of the A12-34 split while Lue Jinghong would be grouped with Shan on the basis of the A123-4 split. This raises interesting questions which are beyond the scope of the present work. For the present purpose however it is sufficient to say that based on phonological evidence, Lue Murng Yorhng is closer to Khuen varieties than Lue Jinghong. Moreover linguistic features do not necessarily align with ethnic group boundaries so it is quite possible that two Lue varieties can legitimately be classified in different subgroups of SWT. Moerman (1965:1217-18) emphasised the diversity of Lue varieties.

> There is sometimes as much apparent speech divergence between the Lue (or Yuan [Archer 1888:13]) of different districts as between a variety of the Lue dialect and some other, non-Lue, dialect. Ban Ping[40] informants, for example, report that the Khyn of Chiengtung "speak as we do," but claim that the Lue of

[40] Chiang Kham District, Pha Yao Province, North Thailand. Chiang Kham was formerly part of Chiang Rai Province but in 1977 Pha Yao was separated from Chiang Rai as an administrative entity.

Chiengkhawng are very difficult to understand. Khyn as described by Egerød (1959; 1961:49-58) and White Thai as described by Minot (1940) both appear quite similar to the Phong variety of Lue spoken in Ban Ping.

The entries in Table 68 show those PT initial consonants for which there is variation in the modern reflexes among the selected varieties of interest in this thesis. First note that Shan and Mao are identical with respect to these consonants. Also note that whereas the Lue varieties pattern with Northern Thai for the phonemes /b/, /f/ and /d/ they pattern with Shan and Mao for the phonemes /j/ and /k^h/. Lue Murng Yorhng is like Northern Thai in that it retains the aspirated palatal affricate /c^h/ albeit only in some varieties, whereas Lue Jinghong is like Shan and Mao in having only its unaspirated counterpart /c/.

PT	Standard Thai	Northern Thai	Lue Jinghong	Lue Murng Yorhng	Shan	Mao
***ʔb**	b	b	b	b	m	m
***f**	f	f	f	f	p^h	p^h
***v**	f	f	f	f	p^h	p^h
***ʔd**	d	d	d	d	l	l
***ch**	c^h	(c^h)[41] (c)	c	c^h	c	c
***ɲ**	j	ɲ	j	j	j	j
***hɲ**	j	ɲ	j	j	j	j
***kh**	k^h	k^h	x	x	k^h	k^h
***x**	k^h	k^h	x	x	k^h	k^h
***ɣ**	k^h	k^h	x	x	k^h	k^h

Table 68 Initial consonant differences between selected Tai varieties

The motivation for examining these phonological features is to see whether there are any other features that Lue Murng Yorhng shares with Northern Thai and Khuen that it does not share with Lue Jinghong and Shan. There are no such features in the above table since Khuen does not have the aspirated palatal affricate /c^h/. As far as vowels are concerned, the main differences seem to be that

[41] Some varieties have /c^h/ whereas others have /c/.

where Northern Thai has diphthongs, Khuen and Lue have single vowel phonemes. There is thus little evidence from consonant or vowel correspondences to further identify which of Lue Murng Yorhng or Northern Thai is the closest to Khuen.

6.6 Conclusions Relating to Goal 4

The conclusions to this section compare the groupings from lexical similarity and phonological features. Standard Thai has the lowest lexical similarity with Khuen followed by Tai Mao. A similar pattern is shown by the classification based on Chamberlain's (1975) and Robinson's (1994) phonological criteria hierarchy, namely that Standard Thai separates off from Khuen at the highest division of SWT (*b>ph; *d>th etc.) and Tai Mao at the next highest branch (*khw>kh; *kw>k; A1-23-4 etc.). Following Robinson's classification Northern Thai and Lue Murng Yorhng are the varieties closest to Khuen but there is no further evidence from phonological segments to establish either of these two as closer than the other. Lexically, both Lue Murng Yorhng and Lue Jinghong are closer than Northern Thai but the differences are so slight as to probably not be of practical importance, despite being statistically significant.

6.7 Evaluation of Survey Instruments

This section makes note of areas for improvement in the survey instruments used on the survey as well as aspects of the survey that seemed to work particularly well.

6.7.1 Evaluation of Khuen 406 Wordlist

Despite the deliberate design of the wordlist to provide coverage of various phonological features, there were still categories that were not well covered. This was in part due to the fact that the wordlist was translated from English into Burmese and thence into Shan. The intended cognate was not elicited because the

semantic thread was lost in translation. Certain words were not sufficiently clear for easy elicitation in a field context, for example, item 271 'barbed spike'.

Direct use of PT forms in the evaluation would provide a firmer basis for evaluation. As the findings of the lexicostatistical comparison showed, Standard Thai is the least closely related of the languages studied to Khuen.

6.7.2 Evaluation of sociolinguistic questionnaires

There were several questions on the questionnaires that did not seem to work well. Q61-Q63 on the ISQ relating to Khuen dialect perceptions for example did not provide the kind of data that had been hoped for since people's knowledge of the general Khuen region was limited to the areas immediately surrounding their own village. Q47-Q48 about the most prestigious Khuen variety were also not successful in terms of providing insights into Khuen speakers' attitudes to different Khuen varieties. One weakness was that because of the way the question was translated into Shan, the regional capital, i.e., Keng Tung, was excluded as a possible answer. The translation should therefore be corrected in any further use of this question. In Q47 the vague term 'best' could be replaced by a more specific term such as 'clearest' or 'most beautiful'. This would not necessarily improve the quality of information gained, if as appears to be the case, subjects' knowledge of other varieties is very limited.