CHAPTER 5

LEXICOSTATISTIC ANALYSIS

5.0 Introduction

This chapter illustrates the procedure and the results of the lexical comparison in this study.

5.1 Cognate Decision

When lexical items are compared, only root forms of words with the same gloss are considered. This means presyllables and other extra syllables are eliminated. This is done so related cognate forms can be identified. Extra syllables include grammatical markers as well as derivational or elaborative syllables. Table 14 shows an example of lexical items in full syllable form.

Number	16	26	39	29	40
Gloss	eye	heart	left (side)	know	nose
1. Mangshi	$hui^{11}ta^{33}$	$caɯ^{33}$	$pa^{33}{}'sai^{53}$	hi^{53}	$hua^{35}laŋ^{33}$
2. Ruili	ta^{33}	$caɯ^{33}$	$faj^{11}sai^{53}$	$nəj^{31}hu^{53}$	$hu^{35}laŋ^{33}$
3. Yingjiang	ta^{33}	$caɯ^{33}$	$pa^{31}sai^{53}$	hu^{53}	$hu^{31}laŋ^{33}$
4. Zhefang	ta^{33}	$caɯ^{33}$	$pa^{31}sai^{53}$	$hu^{53}dʒak^{11}$	$hu^{35}laŋ^{33}$
5. Shuangjiang	$mak^{11}ta^{33}$	$ho^{35}ca^{33}$	$kʰɔn^{33}sai^{53}$	hu^{53}	$xun^{35}naŋ^{11}$
6. Gengma	ta^{33}	$caɯ^{33}$	$pa^{11}sai^{53}$	hu^{53}	$laŋ^{33}$
7. Mengting (Shui Tai)	$mak^{11}ta^{35}$	$ho^{35}caɯ^{35}$	$pa^{11}kʰa^{35}$	$hu^{53}han^{35}$	$xon^{11}nəu^{35}$
8. Mengting(Tai Nua)	$mak^{11}ta^{33}$	ca^{33}	$pa^{11}kʰaŋ^{35}$	hu^{53}	$ho^{35}laŋ^{33}$
9. Canyuang	$mak^{11}ta^{33}$	$ho^{35}ca^{33}$	$xan^{31}sai^{53}$	hu^{53}	$hɔŋ^{53}laŋ^{33}$
10. Lincang	$maʔ^{31}ʈta^{33}$	cau^{33}	$faj^{31}sai^{31}$	$tɔŋ^{53}$	$hu^{35}laŋ^{33}$
11. Fengqing	ta^{33}	$caɯ^{33}$	$xuɔn^{35}sai^{53}$	fu^{31}	$kon^{31}laŋ^{33}$
12. Baoshan	$hui^{11}ta^{53}$	$ho^{35}tsaɯ^{33}$	$pa^{31}sai^{31}$	$hu^{53}laj^{31}$	$hu^{31}laŋ^{33}$
13. Jinghong	$mak^{11}ta^{33}$	$ho^{35}cau^{33}$	$pai^{11}sai^{11}$	hu^{11}	$ho^{35}laŋ^{33}$
14. Simao	$ho^{35}ta^{33}$	$ho^{35}cai^{33}$	$xon^{35}sai^{31}$	$hu^{11}ʔa^{31}$	$laŋ^{33}$
15. Jinggu	ta^{53}	$ho^{35}cai^{53}$	$xan^{11}ʃai^{11}$	hu^{11}	$laŋ^{31}$
16. Tai Mau (Nam Kam)	ta^{21}	$ho^{45}cəɯ^{21}$	$pʰai^{21}sai^{31}$	hu^{31}	$hu^{33}laŋ^{21}$
17. Tai Ya	$'ta:^{3}$	$ho^{34}\,'tsaj^{31}$	$xan^{4}\,'sa:j^{43}$	$'hu:^{31}$	$hu^{42}\,'laŋ^{2}$
18. Tai Lai in Khatcho	hʷita	ho səi	pasait	huʔ	kʰunnəŋ
19. Tai Lai in Khatcho Maungkham	hʷita	mək sai	paʃait	huʔ ʊ	kʰunnəŋ
20. Tai Lai in Ketda	ta	mək sau	saik	hu:ʔ	hu nəŋ
21. Tai Lai in Ywatit	hʷita	mək sau	pasait	hu sək/hu:ʔ	hu nəŋ
22. Tai Lai in Homalin	hʷita	mək səu	ginsaik	hu:ʔ	hu nəŋ

Table 14: Examples of full syllable forms

Table 15 shows the root syllables of those examples after the presyllables and other extra syllables are eliminated.

Number	16	26	39	29	40
Gloss	Eye	heart	left (side)	know	nose
1. Mangshi	ta^{33}	cau^{33}	sai^{53}	hi^{53}	$la\eta^{33}$
2. Ruili	ta^{33}	cau^{33}	sai^{53}	hu^{53}	$la\eta^{33}$
3. Yingjiang	ta^{33}	cau^{33}	sai^{53}	hu^{53}	$la\eta^{33}$
4. Zhefang	ta^{33}	cau^{33}	sai^{53}	hu^{53}	$la\eta^{33}$
5. Shuangjiang	ta^{33}	ca^{33}	sai^{53}	hu^{53}	$na\eta^{11}$
6. Gengma	ta^{33}	cau^{33}	sai^{53}	hu^{53}	$la\eta^{33}$
7 Mengting (Shui Tai)	ta^{35}	cau^{35}	k^ha^{35}	hu^{53}	$xon^{11}nau^{35}$
8. Mengting (Tai Nua)	ta^{33}	ca^{33}	$k^ha\eta^{35}$	hu^{53}	$la\eta^{33}$
9. Canyuang	ta^{33}	ca^{33}	sai^{53}	hu^{53}	$la\eta^{33}$
10. Lincang	ta^{33}	cau^{33}	sai^{31}	$to\eta^{53}$	$la\eta^{33}$
11. Fengqing	ta^{33}	cau^{33}	sai^{53}	fu^{31}	$la\eta^{33}$
12. Baoshan	ta^{53}	$tsau^{33}$	sai^{31}	hu^{53}	$la\eta^{33}$
13. Jinghong	ta^{33}	cau^{33}	sai^{11}	hu^{11}	$la\eta^{33}$
14. Simao	ta^{33}	cai^{33}	sai^{31}	hu^{11}	$la\eta^{33}$
15. Jinggu	ta^{53}	cai^{53}	$\int ai^{11}$	hu^{11}	$la\eta^{31}$
16. Tai Mau (Nam Kam)	ta^{21}	$c\partial u^{21}$	sai^{31}	hu^{31}	$la\eta^{21}$
17. Tai Ya	$'ta{:}^3$	$tsaj^{31}$	$sa{:}j^{43}$	$'hu{:}^{31}$	$la\eta^2$
18. Tai Lai in Khatcho	ta	$s\partial i$	$sait$	$hu\textipa{P}$	$n\partial\eta$
19. Tai Lai in Khatcho Maungkham	ta	sai	$\int ait$	$hu\textipa{P} u$	$n\partial\eta$
20. Tai Lai in Ketda	ta	sau	$saik$	$hu{:}\textipa{P}$	$n\partial\eta$
21. Tai Lai in Ywatit	ta	sau	$sait$	$hu{:}\textipa{P}$	$n\partial\eta$
22. Tai Lai in Homalin	ta	$s\partial u$	$saik$	$hu{:}\textipa{P}$	$n\partial\eta$

Table 15: Root syllable forms

Table 15 shows the root syllable forms of the words. These root forms are used in the comparative method. For word 39 'left (side)' Mengting (Shui Tai) and Mengting (Tai Nua) show quite different data from the rest, as does word 40 'nose' in Mengting (Shui Tai). These results are attributed to an incorrectly

elicited item in these two sites. Thus the data for this lexical item for Mengting (Shui Tai) and Mengting (Tai Nua) is excluded from the analysis.

5.2 Wordlists Comparison

In the comparative method, sets of sound correspondences are identified. These sets are assumed to have derived from the same parent phonemes. The approach used for lexical similarity in this analysis focuses on sound correspondences between lexical items in different speech varieties with the same meaning. This approach is not the same as the comparative method but is a rough approximation for determining apparent cognate forms. For example consider the data shown in Table 16. Note that a dash means there is no data available.

Gloss	1. belly, tummy
1. Mangshi	pom^{33}
2. Ruili	pom^{33}
3. Yingjiang	pom^{33}
4. Zhefang	pom^{33}
5. Shuangjiang	$toŋ^{31}$
6. Gengma	$tɔŋ^{31}$
7. Mengting (The North/Shui Tai)	pum^{35}
8. Mengting (The South/Tai Nua)	pum^{33}
9. Canyuang	$tɔŋ^{31}$
10. Lincang	$toŋ^{11}$
11. Fengqing	$tɔŋ^{31}$
12. Baoshan	pom^{33}
13. Jinghong	$tɔŋ^{33}$
14. Simao	$tuaŋ^{31}$
15. Jinggu	$tuaŋ^{31}$
16. Tai Mau (Nam Kam)	pum'^{21}
17. Tai Ya	-
18. Tai Lai in Khatcho	$do^{n}k$
19. Tai Lai in Khatcho Maungkham	$d^{w}oŋ$
20. Tai Lai in Ketda	doŋ
21. Tai Lai in Ywatit	*doŋ*
22. Tai Lai in Homalin	doŋ

Table 16: The data for the word 'belly, tummy' in each location.

When forms from different speech varieties are compared, the phones are classified into one of the categories.

They are compared according to the following Table of criteria for determining lexical similarity as mentioned on chapter 3, Figure 21.

Category 1:   (a) Exact consonant matches,

(b) Vowels or diphthongs differing by 1 or fewer features,

(c) Phonetically similar consonants in 3 or more word pairs, and

(d) A deletion in three or more word pairs.

Category 2:   (a) Phonetically similar consonants in less than 3 word pairs, and

(b) Vowels differing 2 or more features,

Category 3:   (a) Non phonetically similar consonants, and

(b) A correspondence with nothing in less than 3 word pairs.

Ignore (a) A regularly occurring epenthesis.

In Table 17, Mangshi is compared with other speech varieties based on the criteria for determining lexical similarity.

Note that a dash means there is no data available.

Gloss	1. belly, tummy
1. Mangshi-2. Ruili	1a,1a,1a
1. Mangshi -3. Yingjiang	1a,1a,1a
1. Mangshi -4. Zhefang	1a,1a,1a
1. Mangshi- 5. Shuangjiang	3a,1a,3a
1. Mangshi- 6. Gengma	3a,2b,3a
1. Mangshi -7. Mengting (The North/ Shui Tai)	1a,1b,1a
1. Mangshi -8. Mengting (The South/ Tai Nua)	1a,1b,1a
1. Mangshi -9. Canyuang	3a,2b,3a
1. Mangshi -10. Lincang	3a,1a,3a
1. Mangshi -11. Fengqing	3a,2b,3a
1. Mangshi -12. Baoshan	1a,1a,1a
1. Mangshi -13. Jinghong	3a,2b,3a
1. Mangshi -14. Simao	3a,2b,3a
1. Mangshi -15. Jinggu	3a,2b,3a
1. Mangshi -16. Tai Mau (Nam Kam)	1a,1b,1a
1. Mangshi- 17. Tai Ya	-------
1. Mangshi- 18. Tai Lai in Khatcho	2a,1a,2a
1. Mangshi -19. Tai Lai in Khatcho Maungkham	2a,1a,2a
1. Mangshi -20. Tai Lai in Ketda	2a,1a,2a
1. Mangshi-21. Tai Lai in Ywatit	2a,1a,2a
1. Mangshi -22. Tai Lai in Homalin	2a,1a,2a

Table 17: The word 'belly, tummy' in Mangshi is compared with other sites.

Once these categories are determined, other criteria based on word length (number of phones) are used to determine if the words are lexically similar or not. In order to establish whether the words are lexically similar, they must meet the conditions shown in Chapter 3, Figure 22. For example, according to word length shown in Table 18 if there are three phones in one pair of words, then either all the phones must be in category 1 or two of the phones must be in category 1 and one phone in category 2 to be considered as lexically similar.

Word Length		Category 1	Category 2	Category 3
2	=	2	0	0
3	=	2	1	0

Table 18: Word length

For the first three pairs shown in Table 17, there are three phones in category 1 so they are lexically similar with one another. In the fourth pair, there are two phones in category 3 and one 'phone in category 1, so it does not meet the criteria for lexical similarity so they are not considered lexically similar.

Gloss   1. belly, tummy

var 1-2	1
var 1-3	1
var 1-4	1
var 1-5	0
var 1-6	0
var 1-7	1
var 1-8	1
var 1-9	0
var 1-10	0
var 1-11	0
var 1-12	1
var 1-13	0
var 1-14	0
var 1-15	0
var 1-16	1
var 1-17	no comp
var 1-18	0
var 1-19	0
var 1-20	0
var 1-21	0
var 1-22	0

Table 19: The results of the lexical comparison of the word "belly, tummy"

Note: Num 1 = Lexical similarity

Num 0 = Not similar

No comp = No comparison because there is no data available

Table 19 shows the results of the lexical comparison for the word "belly". This analysis shows that the Tai Nua speech varieties spoken in Mangshi, Ruili, Yingjiang, and Zhefang and 3 other locations considered are lexically similar, while the Tai Nua speech variety spoken in Shuangjiang and 12 others is not

considered lexically similar. This lexical comparison process was repeated for each of the 100 lexical items.

The same steps are applied with other speech varieties one by one. Therefore, the speech variety spoken in Ruili is compared with Yingjiang, Zhefang and so forth. Every one of them follows the same method.

5.3 Lexicostatistics Similarity and the analysis of results

After the lexical analysis is done, the Excel program was employed to calculate the percentages of the lexical similarity between each dialect. They were then put into a matrix (Table 40). Based on this matrix, two tree diagrams (Figures 41 and 42) were constructed.

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10	Var 11	Var 12	Var 13	Var 14	Var 15	Var 16	Var 17	Var 18	Var 19	Var 20	Var 21	Var 22
Var 1																						
Var 2	95																					
Var 3	95	93																				
Var 4	**_96_**	94	92																			
Var 5	86	85	84	86																		
Var 6	90	88	88	89	91																	
Var 7	91	89	88	90	86	91																
Var 8	88	85	86	87	87	91	90															
Var 9	83	81	81	82	90	87	83	89														
Var 10	82	81	79	82	87	88	82	80	84													
Var 11	82	80	79	82	85	88	83	81	83	84												
Var 12	95	94	91	94	87	89	90	86	82	81	83											
Var 13	82	80	80	82	90	89	81	83	83	85	81	82										
Var 14	80	80	78	79	82	83	80	77	79	83	78	81	86									
Var 15	87	86	84	87	88	89	86	84	82	86	82	87	90	89								
Var 16	92	92	91	91	82	86	88	87	81	77	78	90	78	75	82							
Var 17	81	78	77	81	78	82	81	80	78	80	77	80	82	81	85	82						
Var 18	85	86	85	86	83	84	82	83	81	76	77	84	77	**_74_**	81	88	82					
Var 19	86	87	86	87	82	85	83	84	81	77	78	85	78	75	83	89	80	94				
Var 20	83	84	83	84	81	82	80	81	80	**_74_**	76	84	76	75	80	86	78	90	92			
Var 21	84	86	84	84	81	83	81	81	80	76	77	85	76	76	81	88	78	90	93	95		
Var 22	83	84	83	84	81	82	80	81	80	76	76	82	77	76	81	87	80	90	94	93	99	

Figure 40: A Matrix of lexical similarity percentages

Note: Var 1 = Mangshi, Var 2 = Ruili, Var 3= Yingjiang, Var 4 = Zhefang, Var 5 = Shaungjiang, Var 6 = Gengma, Var 7 = Mengting the North (Shui Tai), Var 8 =

Mengting the South (Tai Nua), Var 9 = Canyuan, Var 10 = Lincang, Var 11 = Fengqing, Var 12 = Baoshan, Var 13 = Jinghong, Var 14 = Simao, Var 15 = Jinggu, Var 16 = Tai Mao in Namkham , Var 17 = Tai Ya, Var 18 = Tai Lai in Khatcho, Var 19 = Tai Lai in Maungkham, Var 20 = Tai Lai in Ketda, Var 21 = Tai Lai in Ywait, Var 22 = Tai Lai in Homalin.

This analysis focuses on only Tai Nua in Yunnan, China. According to the matrix in Figure 40, Tai Nua speech varieties are generally quite similar. The minimum lexical similarity is 77 percent between Simao and Mengting. Speech varieties with less than 80 percent lexical similarity mostly occur between Tai Nua speech varieties in Ying Jiang and Simao. This is probably because geographically Simao is quite isolated and is very distant from Ying Jiang (see Figure 39). The highest lexical similarity between Tai Nua speech varieties is 96 percent. It is a pair between Tai Nua spoken Mangshi and Zhefang. This high percentage is not surprising since Zhefang is quite close to Mangshi. (Figure 39 is the map of selected sites).

5.4 Tree Diagram

The "Unweighed Pairs Grouped Method with Arithmetic Average" method, or UPGMA, and programs in the PHYLIP 3.6 suite (Feldsenstein 2002) are used to construct the following non-rooted tree diagram. A non-rooted tree is less constrained than the typical rooted tree. The length of lines on this tree shows how different the varieties are, thus the longer the line the more different, etc. This tree diagram helps to see which dialects can be grouped in one grouped together.
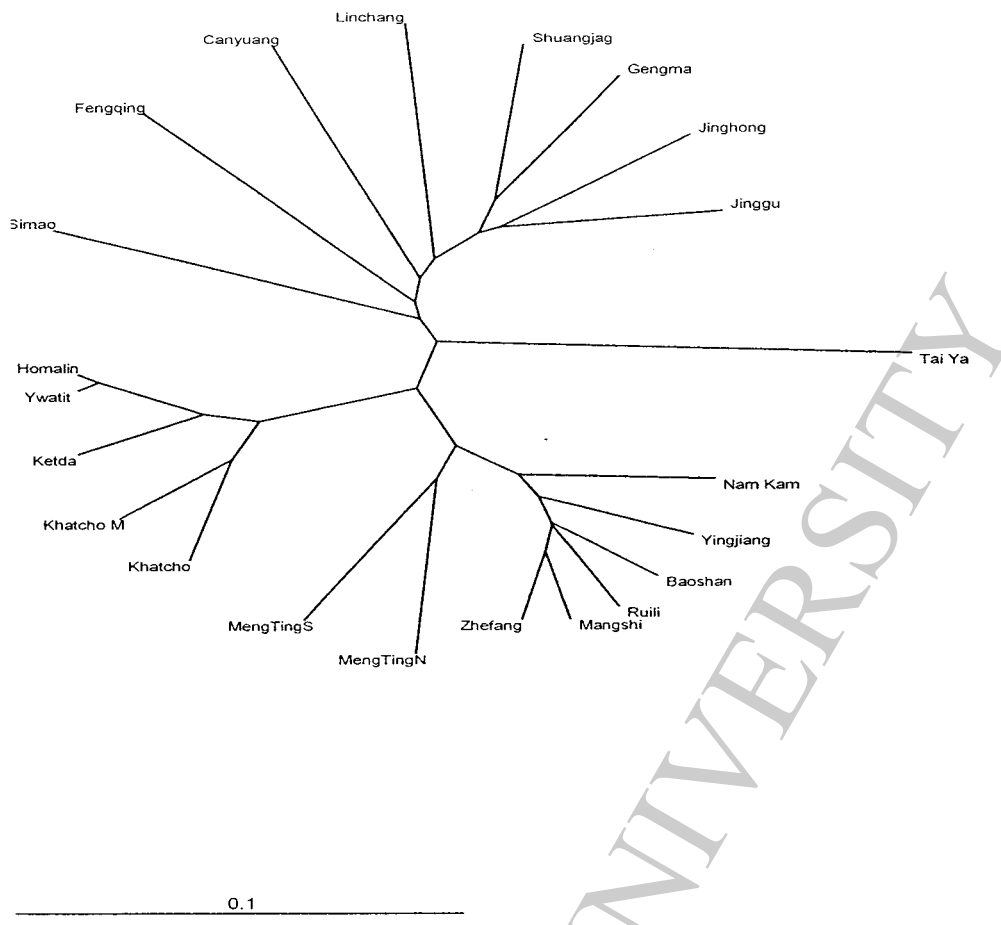
Figure 41: Unrooted tree for Tai Nua, Tai Mau and Tai Lai varieties

According to the Figure above, the speech varieties are roughly grouped into four main groups. These are

1) Tai Mao in Nam Kham, Tai Nua in Baoshan, Mangshi, Ying Jiang, Zhefang, Ruili, Mengting in the North, Mengting in the south

2) Tai Lai in Khatcho, Khatcho Maungkham, Ketda, Ywait, and Homalin

3) Tai Nua in Shuang Jiang, Gengma, Jinghong, Jinggu, Lincang, Fengqing, Canyuan, and Simao

4) Tai Ya

The length of the lines indicates different degrees of lexical similarity. The shorter the line is, the more similar the speech varieties are. Therefore, the degree of the lexical similarity is found to be different in each group. The speech varieties in group 1 share more similarity with each other than those in group 3.

According to the line's length, in group 1 they could be divided into two minor Tai Nua groups:

1.1) Tai Nua in Zhefang, Mangshi, Ruili, Baoshan, Yingjiang and Namkham, and

1.2) Tai Nua in Mingting the North and Tai Lue in Mengting the South.

Tai Nua in Zhefang is the closest to Mangshi, followed by Ruili, Baoshan, Yingjiang and Namkham.

In group 2, all speech varieties under this group are recognized as Tai Lai in Burma. The split between Tai Lai in Homalin and Ywait is very small, indicating that lexical similarity between them is very close while Tai Lai in Khatcho M is closer to Khatcho than Ketda.

In group 3, the speech varieties of Tai Nua which are found under this group, are located in the lower areas in Sipsongbanna. Geographically, Simao is quite separate so the lexical similarity is shared the least with others. Two pairs are found in this group: Tai Nua in Shuangjiang and Gengma, and Tai Nua in Jinghong and Jinggu.

Another picture illustrating the speech varieties divisions is presented in Figure 42. This is a rooted tree constructed using the UPGMA (J. Grimes 1995:69) method and programs in the PHYLIP 3.6 suite (Feldsenstein 2002). In this Figure, the speech varieties in group 1 and 2 are under the same branch while group 3 is under another. This is interesting because the speech varieties in group 1 and 3 are reported as Tai Nua while group 2 is recognized as Tai Lai. This would seem to indicate that Tai Lai is lexically an intermediate branch between two Tai Nua groups. Further research will be needed to clarify the relationship between Tai Nua and Tai Lai.

Besides, Figure 42 clearly presents 6 pairs of some speech varieties showing that they are closer to each other than others. They are:

1) Tai Nua in Mangshi and Zhefang,

2) Tai Nua in Mengting the North and Tai Lui in Mengting the south,

3) Tai Lai in Khatcho M and Khatcho,

4) Tai Lai in Ywait and Homalin,

5) Tai Nua in Shaungjiang and Gengma, and
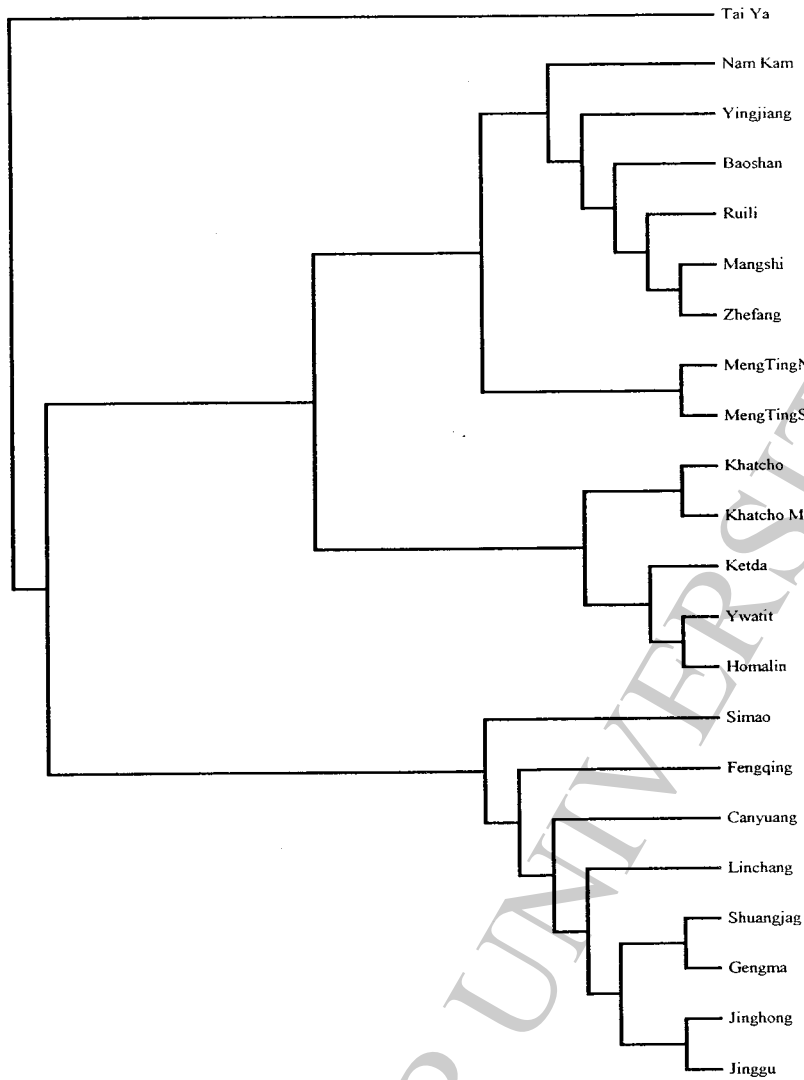
6) Tai Nua in Jinghong and Jinggu.

Figure 42: Rooted tree for Tai Nua, Tai Mau and Tai Lai varieties.

5.5 Chaining and Dominance patterns

In this thesis, two common patterns of convergence involving more than three speech varieties are discussed. These two patterns are chaining and dominance.

5.5.1 The chaining pattern

Chaining occurs when one speech variety is lexically most similar to the neighboring speech variety in a chain. This is likely due to contact. Consider the pattern shown in Figure 30. Likewise, the speech varieties of Tai Nua further away from each other show lower lexical similarity. The distance between each of the selected sites show the chaining pattern in 1)Yingjiang (YJ), 2) Mangshi (MS), 3) Gengma (GM), and 4) Simao (SM). Figure 43 below shows geography of Tai Nua distribution in Ying Jiang, Mangshi, Gengma, and Simao.
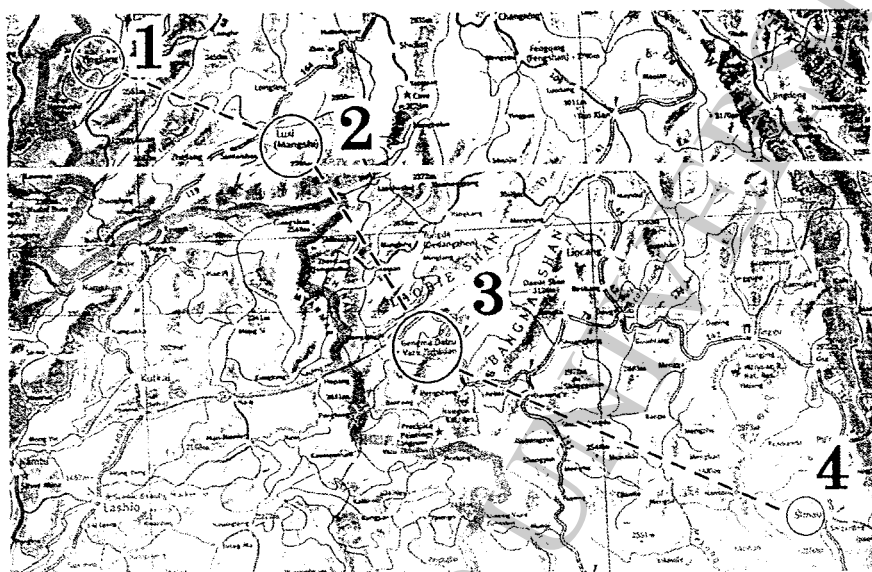

Figure 43: Map showing Tai Nua distributions in 4 locations

When the lexical similarity is presented in a matrix, the highest percentages are customarily shown on the diagonal and the lowest percentages are shown on the lower left hand corner as in Figure 31 in chapter 3. Likewise in Figure 44, the percentages in the left side are lower than the right side.

YJ

95	MS		
88	90	GM	
78	80	83	SM

Figure 44: Cognate percentages

Figure 44 shows the lexical similarity between Ying Jiang and Mangshi is 95 percent, Ying Jiang and Gengma is 88 percent and Ying Jiang and Simao is 78 percent. Figure 44 shows that the lexical percentage decreases as the geographical distance between speech varieties increases.

5.5.2 The dominance pattern

Dominance is when one speech variety, which is geographically central, has the highest lexical percentages with the other speech varieties. Like chaining, this is also due to contact.

The map in Figure 45 shows that 1) Gengma (GM) in the center surrounded by 2) Mengting (MT), 3) Lincang (LC), 4) Shuangjinag (SJ), and 5) Canyuan (CY)
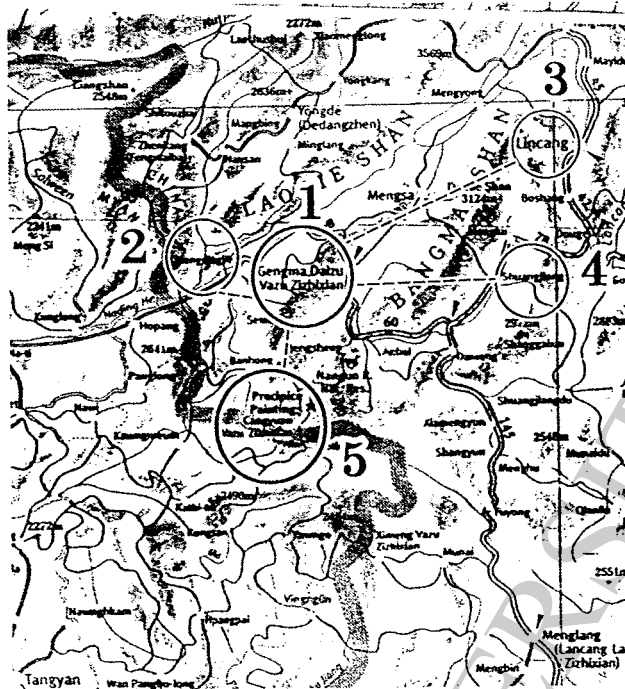
Figure 45: Map showing the location of Gengma, Mengting, Lincang, Shuangjiang and Canyuan.

The cognate percentages are shown in Figure 46:

GM

91	MT			
89	83	CY		
91	87	87	SJ	
88	80	80	87	LC

Figure 46: Cognate percentages

Figure 46 shows that Gengma shares greater lexical similarity with Mengting, Canyuan, Shuangjiang, and Lincang than the latter speech varieties share with one another. For example, the lexical similarity between Gengma and Canyuan is 89 percent while Mengting and Canyuan is 83 percent.

This chapter displays the results of lexical comparison of all selected speech varieties from lexicostatistic analysis. It is found that the Tai Nua speech varieties from Mangshi and Zhefang share the highest lexical similarity (96%) while Mengting in the South and Simao share the least (77%). The rooted and unrooted tree diagrams illustrate that that the speech varieties are basically divided into 4 groups. These results correlate well with geographic proximity as the closest cities tend to share higher lexical similarity.