

CHAPTER 2

SELECTION OF REPRESENTATIVE CHIN LANGUAGES

2.0 Introduction

This section discusses how representative languages were selected out of the 21 Chin languages under study for comparative analysis. Lexicostatistic methods were applied to all 21 languages. Based on these results a preliminary subgrouping was proposed, and then representative languages were selected from each subgroup.

2.1 Lexicostatistics analysis

Lexicostatistics as a technique is generally associated with glottochronology, the attempt to date the division of languages and dialects from each other on the basis of lexical divergence. Crystal (1980:221) even defines as a technique “to make quantitative comparisons between the rates of change within sets of lexical items in hypothetically related languages ...”. However, in the present work lexicostatistic comparison are used to characterize the general degree of divergence of the languages from each other so that representative languages may be chosen as the basis for historical reconstruction.

Lexicostatistic methods were applied to 21 different Chin languages spoken in Myanmar. The 100 cognates (89 Swadesh and some others from the SIL MSEA wordlist) were chosen and compared between the Chin varieties to determine the degree of lexical similarity. The lexicostatistic method used in this study first determines the root word, assuming that the proto language is monosyllabic (Matisoff 1973). Suspected borrowed words were screened out (see more section 4.1). Then the possible morphological markers and non-root syllables were ignored. Pairs of roots

are then compared on a phoneme-by-phoneme basis, with each pair of phonemes assigned to one of 3 categories according to the following criteria (Blair 1990).

- Category 1. (a) Exact matches.
 (b) Vowels or diphthongs differing by one feature.
 (c) Phonetically similar segments in three or more word pairs.
- Category 2. (a) Phonetically similar segments in fewer than three word pairs.
 (b) Vowels or diphthongs differing by two or more features.
- Category 3. (a) Non-phonetically similar segments.
 (b) A correspondence with nothing in fewer than three word pairs.
- Ignore (a) Reduced syllables and non-root syllables.
 (b) A regularly occurring deletion.
 (c) Tone.

Categories 1(c) and 2(a) relate to the frequency of occurrence of phonetically similar segments. 1(c) is assumed to show a higher degree of probability than 2(a) that the correspondence is not random. Categories 1(b) and 2(b) assume that vowels or diphthongs differing by one feature have a higher degree of similarity than those differing by two or more features.

After assigning the categories for each phone correspondence, the category assignments for each word were tabulated. Then Table 2 adapted from Blair (1990:32) was used to determine whether the pair of words was sufficiently similar to be considered “lexically similar”.

Phonemes	Category 1	Category 2	Category 3
2	2	0	0
3	2	1	0
4	2	1	1
5	3	1	1
6	3	2	1
7	4	2	1

Table 2. Phone table for lexical similarity (Blair 1990:32)

After each pair of languages had been compared, a percentage of lexical similarity was calculated. The lexicostatistic similarity matrix for the 21 Chin languages is given in Table 3. Language names are arranged in approximate north-south order. It is interesting to note that the percentages of lexicostatistic similarity among the northern languages are higher than among the southern languages.

	Zo	Sinyin	Tedim	Bual	Zanniet	Mizo	Falam	Taisun	Hakha	Thantlang	Khualsim	Senthang	Matu	Kaang	Dai	Asho	Lautu-H	Lakher	Mara	Khumi
Thado	82	81	80	67	67	59	65	62	62	62	60	50	47	48	47	50	39	40	38	40
Zo		85	88	67	66	59	63	61	61	61	61	50	43	45	41	47	38	33	34	38
Sinyin			91	64	69	62	69	65	62	62	62	53	45	46	43	49	36	35	33	37
Tedim				66	69	63	66	63	67	65	63	52	44	46	47	50	39	35	34	38
Bualkhu					88	65	69	69	64	63	60	53	43	45	43	44	37	33	28	33
Zanniet						68	75	75	71	70	68	66	44	47	45	49	41	34	34	38
Mizo							72	76	73	73	65	55	42	47	33	47	44	36	32	31
Falam								87	82	81	80	59	46	43	40	44	38	35	33	34
Taisun									83	81	77	62	44	44	38	41	39	36	32	34
Hakha										95	89	67	47	47	42	47	46	38	36	37
Thantlang											76	66	48	46	43	47	49	35	38	36
Khualsim												68	43	41	37	46	41	38	37	36
Senthang													50	47	41	42	45	38	37	38
Matu														38	35	42	47	38	35	37
Kaang															58	42	37	33	30	48
Dai																48	39	32	31	31
Asho																	36	28	21	36
Lautu-H																		52	51	34
Lakher																			81	34
Mara																				33

Table 3. Matrix of lexicostatistic percentages in 21 Chin language

The percentages of lexicostatistic similarity in Table 3 are consistent with Luce's (1985) findings in which Thado (a northern language) has a 87.8% as opposed to Khumi (a southern language) which has only a 60.4% similarity with other Chin.

2.2 Turning numbers into trees

In order to better visualize the lexical similarity relationships between these 21 languages, the matrix in Table 3 can be transformed into a tree diagram. The "Unweighed Pairs Grouped Method with Arithmetic Average" (UPGMA, or 'Average Link') is a method for doing this that makes minimal assumptions about the data (Grimes 1995: Appendix 1). The table of apparent cognate percentages was processed using the "Cluster Analysis 1.01" computer program (Quigly 1995), which implements the UPGMA method¹¹.

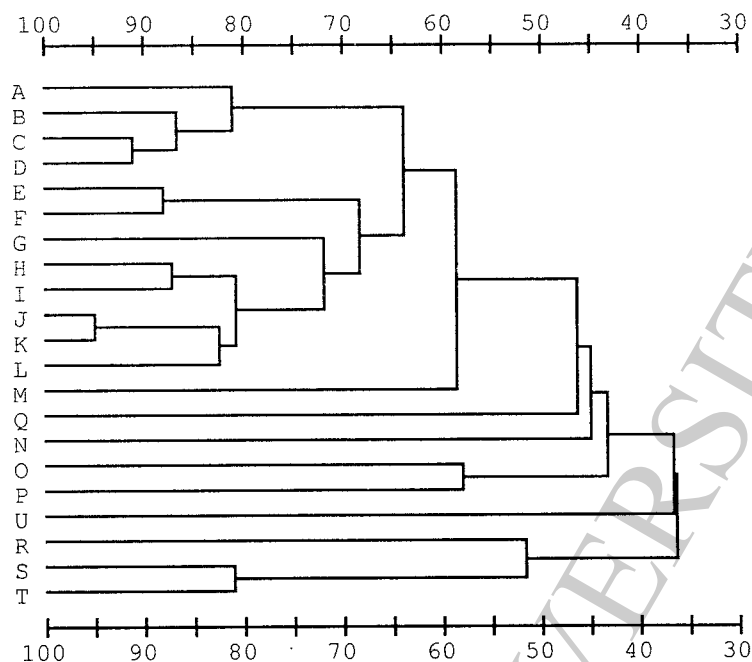
Table 4 presents the levels of average similarity at which each language is related to others in the tree structure.

¹¹The Cluster Analysis 1.01 program was run by Dr. J. F. Bennett.

Percentage of similarity	Names of languages
95.0	Hakha, Thantlang
91.0	Siyin, Tedim
88.0	Bualkhua, Zaniat
87.0	Falam, Taisun
86.5	Zo, Siyin, Tedim
82.5	Hakha, Thantlang, Khualsim
81.0	Thado, Zo, Siyin, Tedim
81.0	Lakher, Mara
80.7	Falam, Taisun, Hakha, Thantlang, Khualsim
71.8	Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim
68.1	Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim
63	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim
58.4	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang
58.0	Kaang, Dai
51.5	Lautu-Hnaring, Lakher, Mara
46.4	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang, Asho
44.9	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang, Asho, Matu
43.2	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang, Asho, Matu, Kaang, Dai
36.6	Thado, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang, Asho, Matu, Kaang, Dai, Khumi
36.1	Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thantlang, Khualsim, Senthang, Asho, Matu, Kaang, Dai, Khumi, Lautu-Hnaring, Lakher, Mara

Table 4. Percentage of lexicostatistics similarity

Figure 13 presents the same information as a tree diagram.



Code	Language name	Code	Language name	Code	Language name
A	Thado	H	Falam	O	Kaang
B	Zo	I	Taisun	P	Dai
C	Siyin	J	Hakha	Q	Asho
D	Tedim	K	Thantlang	R	Lautu-Hnaring,
E	Bualkhua	L	Khualsim	S	Lakher
F	Zaniat	M	Senthang	T	Mara
G	Mizo	N	Matu	U	Khumi

Figure 1. Chin language tree based on average link method

The correlation between the grouping in Table 4 and the original data (i.e. Table 3) is 0.968. That is, the process of transforming Table 3 to Table 4 distorts the data by only a very small amount.

2.3 Preliminary subgrouping

The preliminary subgrouping shown in Figure 14 is based on standard lexicostatistic procedures. (Figure 14 re-casts Figure 13 into a more familiar stammbaum format). Percentages of lexicostatistic similarity show that there are two main Chin language groups: a Northern group consisting of those languages below and to the left of the

dotted line (traditionally the Northern and Central Chin languages), and a Southern group. The Northern group forms a relatively tight cluster, with lexicostatistic similarity counts of 63% and above; the Southern group shows much more internal diversity.

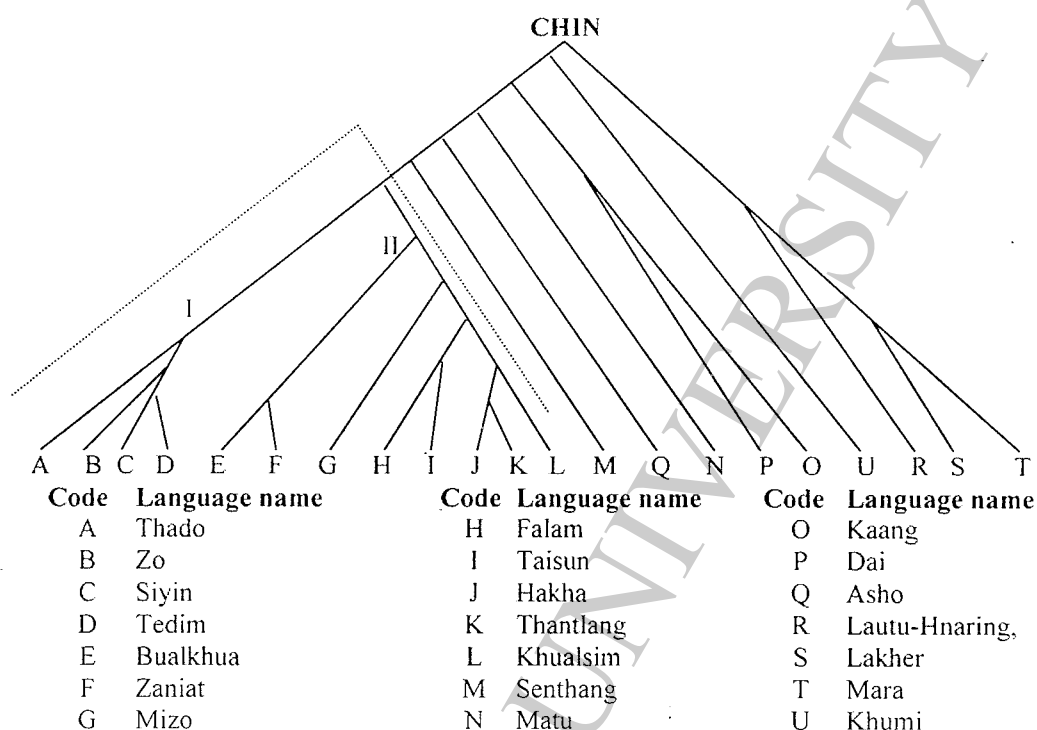


Figure 14. Preliminary subgrouping of Chin languages

The Northern languages (Thado, Zo, Siyin, Tedim, Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thangtlang and Khualsim) can be subdivided into two subgroups: I (Thado, Zo, Siyin and Tedim) and II (Bualkhua, Zaniat, Mizo, Falam, Taisun, Hakha, Thangtlang and Khualsim).

Clear subgrouping of the southern languages is difficult. Based on the tree above, it is clear that Lautu-Hnaring, Lakher and Mara are in one group and Dai and Kaang are in one group. According to So-Hartman (1988) Matu is clustered together with Dai and Kaang under Cho subgroup (see Figure 11).

Therefore there may be at least three subgroups among Southern languages: Lautu-Hnaring, Lakher and Mara in one group; Matu, Dai and Kaang in another; and Khumi in a third. Asho can be clustered with Dai and Kaang group. Senthang is closer to Subgroup II of the Northern language family.

2.4 Selected languages

The preliminary subgrouping of 5 subgroups provides a useful criterion for the selection of representative languages. Six languages were selected for a phonological reconstruction of Proto Chin, one from each subgroup. As subgroup II contains almost half of the languages in this analysis, two languages were selected, Mizo and Hakha. Hakha was chosen on the account of being the dominant language in the central part of the Chin State and having a high degree of similarity with languages within the same group. On the other hand, Mizo was chosen as it has been comparatively more studied by linguists, (although the majority of speakers live in Mizoram State of India). The selected languages are Tedim, Mizo, Hakha, Mara, Khumi, and Kaang, shown in bold type in Table 5, which is based on the above discussion of preliminary subgrouping.

Preliminary Subgrouping of Chin languages				
A		B		
I	II	III	IV	V
A. Thado	E. Bualkhua	N. Matu	R. Lautu	U. Khumi
B. Zo	F. Zaniat	O. Kaang	S. Lakher	
C. Siyin	G. Mizo	P. Dai	T. Mara	
D. Tedim	H. Falam	Q. Asho		
	I. Taisun			
	J. Hakha			
	K. Thantlang			
	L. Khualsim			
	M. Senthang			

Table 5. Selected Chin languages

Tedim represents Group I. Mizo and Hakha represent Group II. Kaang represents Group III. Mara represents Group IV and Khumi represents Group V.

1050001575

Geographically, the representative languages are well distributed in the Chin State. Tedim live in the extreme north of Chin State. The Mizo¹² live in south-western Tedim and western Falam Townships. The Hakha and Mara speaking communities live in the central part of the Chin State. The Kaang live in the southeast of the Chin State in Mindat and Kanpetlet Townships. The Khumi live in Paletwa Township in the southwest of the Chin State close to Bangladesh.

There are several non-linguistic factors which further support the selection of these as representative languages. The selected languages are more prominent in their history of literacy. Grierson (1904:666) states that Khumi orthography was first introduced in 1850 by Rev. L. Stilson¹³. The orthography of Mara (Lakher) was developed by Captain S. R. Tickell in 1852, Siyin by Captain Rundall in 1891, and Hakha by Sergeant-Major A. G. E. Newland (IMS) in 1894. The public use of Tedim orthography was relatively advanced such that beginning in the year 1919, a monthly newspaper in Tedim, the “TEDIM KAM THU KIZAKNA,” was regularly published from Madras until 1938. Chin vernacular education in Tedim, Falam, Hakha and Chinbok was started by the British government in 1925.

Among Chin people, the publication of the Bible in various languages proves significant in the history of literature because “the church is the unofficial ‘keeper of the languages’” (Chhangte 1993:28). Table 6 lists some of the Chin languages (both in India and Myanmar) which have received the Bible (Khup Za Go 1996).

¹² The Mizo living in Tedim and Falam township are known as Hualngo and the majority live in the Mizoram State of India.

¹³ A reader and spelling book were printed but were largely left unused as the missions withdrew.

Language	New Testament	Entire Bible
Mara (Lakher)	1928	1956
Mizo (Lusei)	1917	1959
Hmar	1946	1968
Kuki/Thado	1942	1971
Paite	1951	1971
Tedim	1932	1977
Haka	1940	1978
Vaiphei	1957	1979
Falam	1952	1991
Asho	1954	-
Khumi	1957	-

Table 6. The Bible in Chin languages

Additionally, Mizo (the Duhlian variety) has been used as the lingua franca for the related languages of Chawhte, Hmar, Hnamte, Khawlhing, Khiangte, Ngente, Paihte, Pautu, Pawi, Ralte, Rawite, Renthlei, Tlau, Vangchhia and Zawngte for more than a century (Chhangte 1993:1). Hakha and Tedim are also lingua francas among the languages in their area.

The list of representative languages has the added advantage of allowing reference to previous scholarship. Ono (1965) analyzes Tedim, Ngawn, Hakha, Falam, Anal, Zotung, Khumi and Chinbok. Solnit (1979) uses Tedim and Lushai data, and Bhaskararao (1996) compares the lexicon in Tedim and Lushai. Mizo is probably the most analyzed Chin language.