

CHAPTER 2

LEXICAL COMPARISON

2.0 Introduction

Chapter 1 gave a brief background, overview of each language group, communication, previous research, purpose of this thesis and methodology used/employed. This next chapter will focus on a lexical comparison of the four languages. Lexicostatistic methods were applied to the four languages to give a lexical similarity count and to indicate the lexical relationships between the languages.

2.1 Lexicostatistic Analysis

Bradley (1997) classifies the four languages described in this thesis as Central Karenic. This chapter studies the relationships of these four languages on the basis of lexical divergence. Lexicostatistic methods were applied to Kayah, Kayaw, Monumanaw and Yintale. One hundred words were used in this comparison. These words (listed in Appendix D) are based on the Swadesh 100 with some substitutions appropriate for Southeast Asian languages. These items were compared to determine the degree of lexical similarity. As Karenic languages are primarily isolating, only the root syllable was considered in this comparison. (See Matisoff 1993 for a discussion of Sino-Tibetan syllable structure). Furthermore, Solnit (1997), states that in Kayah most morphemes are monosyllables, though there are a few polysyllables.

Languages were compared word by word for similarity. Each initial consonant element is compared followed by a comparison of each second consonant and then each vowel; morphological markers and non-root syllables were ignored. They were assigned to one of three rankings, based on Table 22.

For each pair of languages a total was calculated by adding up the number of word-pairs that passed the limits set beforehand for determining sufficient similarity. These limits are shown in Table 23. So, for example, a word-pair with three items being compared would need to have at least two rank 1 elements and one rank 2 element to be considered lexically similar.

A similarity percentage was then calculated by dividing the number of apparent cognates with the total number of words compared.

Rank 1	<ul style="list-style-type: none"> a. Exact matches b. Phonetically similar segments in three or more pairs c. Vowels that vary by one feature d. Diphthongs that differ by one feature e. u vs. w, i vs. j f. p vs. b, t vs. d, k vs. g, θ vs. ð g. Segments that are labialized vs. unlabialized h. dz vs. z, ts vs. s, tʃ vs. ʃ, dʒ vs. ʒ i. A deletion that occurs 3 or more times.
Rank 2	<ul style="list-style-type: none"> a. Phonetically similar segments in fewer than three pairs. b. Vowels or diphthongs that differ by two or more features
Rank 3	<ul style="list-style-type: none"> a. Non phonetically similar segments b. A non-regularly occurring deletion
Ignored	<ul style="list-style-type: none"> a. ʔ vs. ø word initial or word final b. Reduced syllables ʔə c. Reduplicated syllables d. Differences due to syllable-level metathesis e. Differences in vowel length f. Syllables that do not appear to be part of the word root

Table 22: Phonetic Similarity Algorithm

Rank 1(b) and 2(a) relate to the frequency of occurrence of phonetically similar segments. 1(b) aims to capture regular sound changes, whereas 2 (a) aims to capture irregular sound changes. Rank 1(c) and 2(b) consider that vowels or diphthongs differing by one feature have a higher degree of similarity than those differing by two or more features.

Phones		Rank 1	Rank 2	Rank 3
1	=	1	0	0
2	=	2	0	0
3	=	2	1	0
4	=	2	1	1
5	=	3	1	1
6	=	3	2	1
7	=	4	2	1
8	=	4	2	2

Table 23. The Phone Table for Lexical Similarity (Blair 1990:32)

Many lexicostatistic comparisons have problems because there is no preset criterion used to establish cognate relationships between word pairs. Two different linguists often reach different conclusions even though they use the same data. Thus, it is important to apply systematic criterion in comparing word pairs. This allows the results to be duplicated.

In the following examples the methodology of determining the ranking correlations for each word are shown:

No	Gloss	Kayah	Monu	R. 1	R. 2	R. 3	Lex. Sim.
1.	Belly	h ə	p ^h u	0	1	1	No
2.	Moon	l ε	l a	2	0	0	Yes
3.	Road	k l j a	k l ø ε	3	0	0	Yes

Figure 12. Examples of Comparison

The phone (Table 23) establishes the minimum conditions that pairs of words must satisfy in order to be considered lexically similar. In figure 12, the first item, the word for *Belly*, has two phones and has a 0-1-1 composite ranking. When compared to table 23, for a word with two phones, the minimum condition for lexical similarity is a 2-0-0. Since this pair of words has a fewer ranking than required for a 2 phone morpheme the minimum conditions are not met. Therefore this pair of words are not lexically similar.

The word *Moon*, has two phones and a 2-0-0 composite ranking. Comparing this with Table 23, the minimum condition for lexical similarity satisfies the condition. Thus these two words are considered lexically similar. For *Road*, with four phones, one of them is ignored, "Ignored a" (the third element [j] in Kayah), so it is considered to be a word of three phones in length. Looking at Table 23, a word with

three phones requires a rank of at least 2-1-0. The word "*road*" is 3-0-0. Since this word pair has more phones in rank 1 than the minimum conditions require, this pair of words are considered lexically similar.

In order to know how closely related the four languages are to each other lexically, the percentage of lexical similarity was calculated for pair-wise comparisons of the languages. The results of the lexical similarity were arranged into a matrix (Table 24).

Kayah			
81%	Yintale		
78%	82%	Monumanaw	
74%	74%	78%	Kayaw

Table 24. Matrix of Lexicostatistic Percentages of Similarity

Table 24 shows the lexicostatistic similarity between Kayaw, Monumanaw, Yintale and Kayah. The lexical similarity percentages linking Kayah and Kayaw is 74%, Kayah and Monumanaw is 78% and Kayah and Yintale is 81%, Kayaw and Monumanaw is 78%, and Kayaw and Yintale is 74%, Monumanaw and Yintale is 82%. Monumanaw and Yintale have higher figures of lexical similarity than other languages.

2.2 Family-tree Relationship

Based on the lexical percentages of these languages, a family tree is shown. This tree depicts how the lexical relationships between these languages can be visualized.

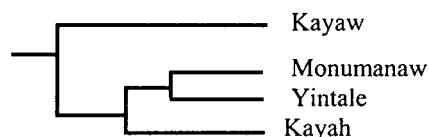


Figure 13. Family – tree Depicting Lexical Relationships

According to Manson (2002:10), Kayah and Yintale are closer to each other than Monu, but the author's analysis is different from what Manson's (2002:18) analysis of the phonological similarity phenogram. The results of the author show that there is a greater difference between Kayaw and the other three, but that there is a greater similarity between Monumanaw and Kayaw. The lexical percentages suggest that Yintale and Monumanaw are more closely related to each other than are the others. One alternative reason why Yintale and Monumanaw are more closely related to each other than the others is that geographically they live very close to each other. According to Matisoff (1978), geographically or culturally contiguous languages may come to have extremely similar phonological inventories by "diffusion," even though they are not genetically related. Kauffman (1993:14) said, "Of all the Central Karen languages, Eastern Kayah has the most contact with Thai. Its phonemic inventory has several similarities to Thai not found in the other Central Karen languages..." However Yintale and Monumanaw do not share any innovative lexical items that no other Karenic language has, which would point to the similarity between them being due to genetic inheritance.

In the following figure of the tree diagram with a scale is shown the lexical similarity relationships between the four languages. It is the same as the above family tree figure, but it is shown with scales for each language in order to be clearer for the reader.

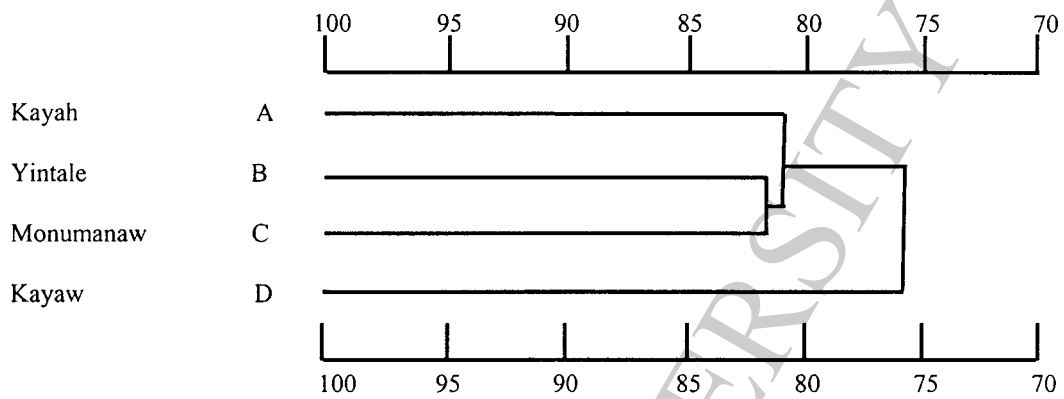


Figure 14. Central Karenic Languages Tree Based on average link method