

CHAPTER 2

LEXICAL COMPARISON

2.0 Introduction

This chapter focuses on the lexical comparison of fifteen speech varieties of Bisoid in order to find the lexical similarity and determine the lexical relationships between these speech varieties.

On the basis of the results of this analysis representative varieties will be chosen (see chapter 2.3) for more in-depth study (Chapters 3 and 4).

2.1 Lexicostatistic Analysis

In defining lexicostatistics, Blair (1990) states that lexicostatistics is the process of comparing suspected cognates in languages to determine the degree of lexical similarity. From the wordlists, the author selected one hundred words following Mann (2004). Each pair of speech varieties was compared word by word to search for degrees of similarity. In this process, the author applied a modified Blair method, as described by Mann (2001), to compare one element to one element of each gloss. In this analysis, only the root elements of the word were considered; prefixes and suffixes were ignored. For the root, the comparison was based on the syllable structure of initial (the consonants) and the rhyme (the vowels). According to Blair (1990), the criterion used to find the lexical similarity can be explained in the following.

Criterion

- Category 1: (a) Exact matches
- (b) Vowel differing by 1 feature
- (c) Phonetically similar segments in 3 or more word pairs
- Category 2: (a) Phonetically similar segments in less than 3 word pairs
- (b) Vowel differing by 2 or more features
- Category 3: (a) None Phonetically similar consonants
- (b) A correspondence with nothing in less than 3 word pairs
- Ignore: (a) Inter-consonantal schwa [ə]
- (b) A regularly occurring deletion

Figure 3. Phonetic similarity according to Blair (1990)

After comparing one element to one element of each cognate, there will be conditions of categories distributed for each element. Therefore, these conditions have to be considered with the phone table in order to consider the lexical similarity. The conditions used to consider can be seen in Table 1.

Number of Phones Compared		Category 1	Category 2	Category 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 1. Phone Table (minimum for lexical similarity)

From table shown above, this table is a list of the minimum score need for words to be considered similar. Longer words have more room for divergence. The reason is to get brief understanding how conditions from comparing cognates work with the phone table.

The methodology of determining the category correlations for each word is explained below.

No	Gloss	Bisu	Tsukong	Ca. 1	Ca. 2	Ca. 3	Lex. Sim.
1.	Speak	c i	c a	1	1	0	No
2.	Dog	k ^h i	k ^h i	2	0	0	Yes
3.	Egg	h j a ? u	? a ŋ ? u 3	3	0	0	Yes

Figure 4. Examples of Comparison

In comparing suspected cognates, words are compared one element by one element basis using the criterion from Figure 3 as seen in Figure. 4 above. Considering the word for *Speak*, there are two phones compared, which are c-c and i-a. The first

phone matches with the category 1(a), which means 'exact matches'. The second phone matches with the category 2(b), which means 'vowels differing 2 or more features'. The vowel /i/ having high, front, and + articulation differs from the vowel /a/ having low, back, and –articulation. The composite ranking is 1-1-0. Looking to the phone table, the composite ranking for the minimum for lexical similarity for two phones is 2-0-0. Comparing to the phone table, it does not match with the minimum ranking. So, this pair of words is not considered lexical similarity.

Considering the word for *dog*, there are also two phones compared, which are k^h-k^h and i-i. The first and the second phone match with the category 1(a), which means 'exact matches'. The composite ranking is 2-0-0. Looking to the phone table, the composite ranking for the minimum for lexical similarity for two phones is 2-0-0. Based on the phone table, it matches with the minimum ranking. So, this pair of words is considered lexical similarity.

Considering the word for *egg*, there are six phones compared, which are h-ʔ, j-ø, a-a, ø-ŋ, ʔ-ʔ and u-u. The first phone matches with the category 2(a), which means 'phonetically similar segments in less than 3 word pairs'. The second phone matches with the ignore (b), which means 'a regularly occurring deletion'. The third phone matches with the category 1(a), which means 'exact matches'. The fourth phone matches with the ignore (b), which means 'a regularly occurring deletion'. The fifth and sixth phones match with the category 1(a), which means 'exact matches'. The composite ranking is 2-0-0. Looking to the phone table, the composite ranking for the minimum for lexical similarity for two phones is 3-0-0. Comparing to the phone table, it matches with the minimum ranking. So, this pair of words is considered lexical similarity.

The lexical relationships of all fifteen speech varieties can be shown in terms of percentages. The percentage of lexical similarity was calculated for pair-wise comparisons of the speech varieties. The results of the lexical similarity are arranged in a matrix in Figure 5.

Coong														
67	Pyen													
75	81	Laomian												
80	63	66	Tsukong											
72	89	90	63	Laopin										
46	39	48	48	37	Mpi									
68	72	80	60	65	39	Phongset								
58	70	72	54	68	38	83	Laoseng							
55	68	79	54	63	42	75	80	Phongku						
51	62	72	52	55	40	80	80	73	Sinsali					
58	57	54	53	59	45	62	57	58	48	Cauho				
65	87	82	63	81	53	68	71	72	65	53	Bisu			
59	54	57	61	61	45	52	50	43	40	51	43	Bantang		
61	80	74	61	79	46	71	69	74	63	52	75	47	Laopan	
54	70	74	53	68	39	76	82	76	78	56	69	46	72	Cantan

Figure 5. Lexicostatistic Similarity of Bisoid Speech Varieties

From all percentages of all fifteen speech varieties, the numbers show that the highest lexical similarity percentages are between Laomian and Laopin with 90%. The lowest percentage of lexical similarity is between Mpi and Laoseng with only 38%.

2.2 Lexical Tree

From the lexical matrix of similarities, a program called Phylip 3.6 generated, using the “Unweighed Pairs Grouped Method with Arithmetic Average” (UPGMA, or Average Link) method, a lexical similarity tree as shown in the following figure.

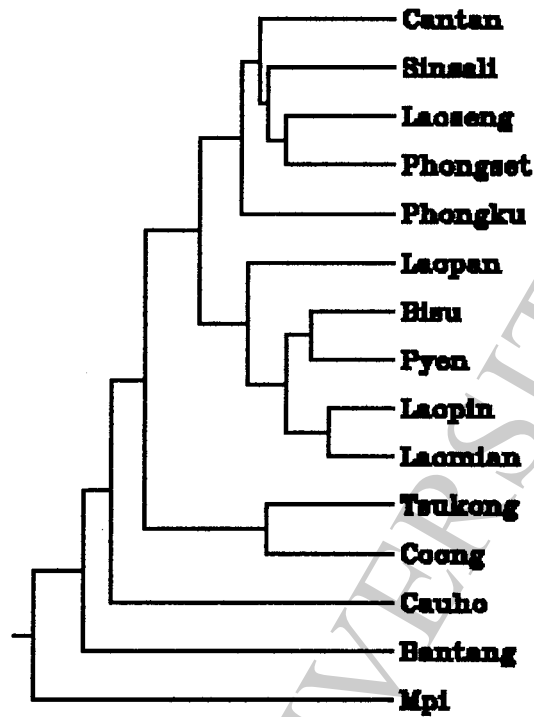


Figure 6. Rooted tree of Bisoid varieties based on UPGMA method

From Figure 6, there are five groups of speech varieties. Group 1 with the most members consists of Cantan, Sinsali, Laoseng, Phongset, Phongku, Laopan, Bisu, Pyen, Laopin, and Laomian. Group 2 consists of Tsukong and Coong. Group 3 consists of Cauho. Group 4 consists of Bantang. Group 5 consists of our Mpioid out-group Mpi.

According to Bradley (1979), Mpi belongs to the subgroup of Mpioid as a different branch from Akoid and Bisoid. From the lexical tree showing, it confirms that Mpi is a separated branch of the subgroup of Southern Loloish. Therefore, this analysis agrees with what Bradley's earlier work.

2.3 Selection of Representative Varieties

From the results of the lexical similarity analysis of the 15 Bisoid speech varieties, a representative variety from each cluster of speech varieties was chosen. As there were four clusters of Bisoid observed, four representatives were chosen. For Group 1 Bisu was chosen as the representative so as to provide a link to previous research. For Group 2 Tsukong was chosen as the data was considered to be more reliable. And for Groups 3 and 4, both which each have only one representative, Cauho and Bantang respectively were chosen.

The lexicostatistic similarity between Bantang, Bisu, Cauho, and Tsukong can be seen in the following table.

Tsukong			
53	Cauho		
63	53	Bisu	
61	51	43	Bantang

Figure 7. Lexicostatistic Similarity of Speech Varieties

The lexical similarity between Bantang and Bisu is 43%, Bantang and Cauho is 51%, Bantang and Tsukong is 61%, Bisu and Cauho is 53%, Bisu and Tsukong, 63%, and Cauho and Tsukong is 53%. Bisu and Tsukong have highest percentages of lexical similarity.