



The Sarawak Museum Journal
Vol. LIII No. 74
December 1998



ISSN: 0375-3050
E-ISSN: 3036-0188

Citation: Paul R. Kroeger. (1998). Language Classification in Sarawak: A Status Report. The Sarawak Museum Journal, LIII (74): 137-174

LANGUAGE CLASSIFICATION IN SARAWAK: A STATUS REPORT*

Paul R. Kroeger

INTRODUCTION

During the past one hundred years and more, wordlists have been published for a large number of the indigenous languages and dialects in Sarawak by anthropologists, missionaries, government servants and travellers. Published accounts are full of vague and impressionistic statements to the effect that "language A is clearly related to language B", or "A and B are merely dialects of the same language", etc. However, little systematic work has been done on the comparison and classification of these languages.

The most significant contributions to our understanding of linguistic relationships in Sarawak are found in the work of Ray (1913), Blust (1974a), Hudson (1978), and Wurm (1983). The present paper is an attempt to summarize and synthesize the results of these and other studies, in the spirit of Cense and Uhlenbeck (1958). The goal is to define what has been accomplished, what remains to be done, and what our priorities should be for further research relating to the classification of Sarawak's indigenous languages. This study cannot claim to be exhaustive; it is intended merely to serve as a catalyst for renewed investigation of this very important topic.

LANGUAGE CLASSIFICATION IN SARAWAK: A STATUS REPORT*

by

Paul R. Kroeger

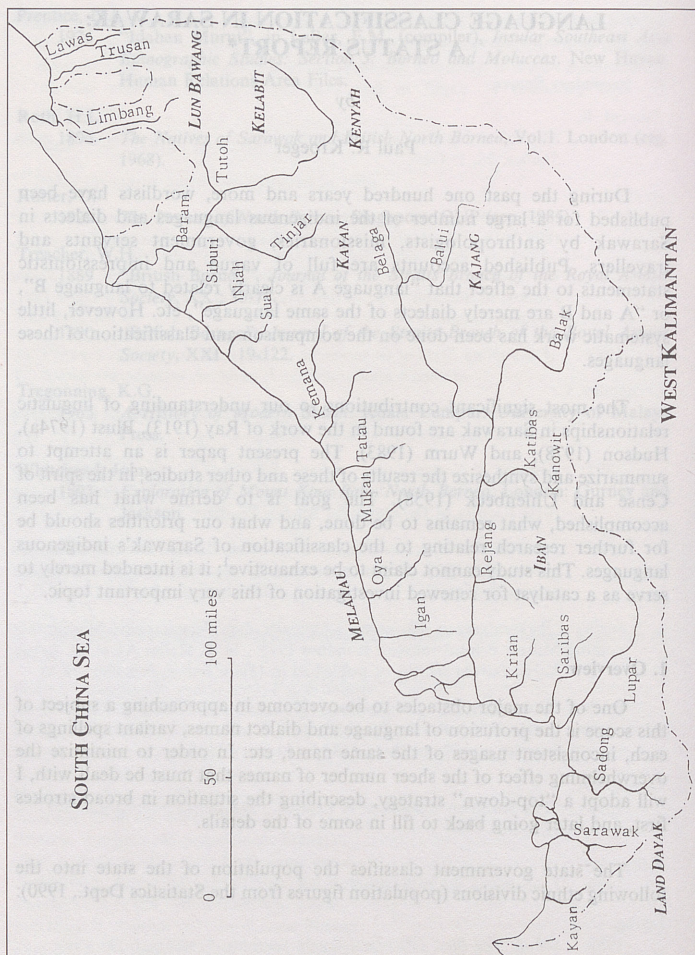
During the past one hundred years and more, wordlists have been published for a large number of the indigenous languages and dialects in Sarawak by anthropologists, missionaries, government servants and travellers. Published accounts are full of vague and impressionistic statements to the effect that "language A is clearly related to language B", or "A and B are merely dialects of the same language", etc. However, little systematic work has been done on the comparison and classification of these languages.

The most significant contributions to our understanding of linguistic relationships in Sarawak are found in the work of Ray (1913), Blust (1974a), Hudson (1978), and Wurm (1983). The present paper is an attempt to summarize and synthesize the results of these and other studies, in the spirit of Cense and Uhlenbeck (1958). The goal is to define what has been accomplished, what remains to be done, and what our priorities should be for further research relating to the classification of Sarawak's indigenous languages. This study cannot claim to be exhaustive¹; it is intended merely to serve as a catalyst for renewed investigation of this very important topic.

1. Overview

One of the major obstacles to be overcome in approaching a subject of this scope is the profusion of language and dialect names, variant spellings of each, inconsistent usages of the same name, etc. In order to minimize the overwhelming effect of the sheer number of names that must be dealt with, I will adopt a "top-down" strategy, describing the situation in broad strokes first, and later going back to fill in some of the details.

The state government classifies the population of the state into the following ethnic divisions (population figures from the Statistics Dept., 1990):



(1)	Ethnic Group	Population	% of total
	Iban	493,000	29.5%
	Chinese	483,000	28.9%
	Malay	348,000	20.8%
	Bidayuh (Land Dayak)	140,000	8.4%
	Melanau	96,000	5.8%
	Other indigenous (= "Orang Ulu")	91,000	5.5%
	Others	19,000	1.1%
	Total:	1,670,000	100%

The IBAN, formerly referred to as Sea Dayak, live primarily in the area between the Rejang and Sadong rivers (see map). They speak a Malayic Dayak language which is more fully documented than any other indigenous language in the state. In addition, the Iban are much more homogenous linguistically than any of the other groups listed in figure (1), apart from the Malays. For this reason, the Iban will receive less attention in the present study than some of the other groups.

The term "Bidayuh" has now replaced the term "Land Dayak" in most contexts, at least in Sarawak. However, the older term still serves a useful purpose as a linguistic classification.² The LAND DAYAK groups live in the interior hill country west of the Sadong river basin. The Land Dayak languages do not appear to be closely related to any other language in Sarawak, but they do form a linguistic subgroup with the many Land Dayak languages spoken across the border in West Kalimantan (Indonesian Borneo).

The MELANAU live along the lower reaches of the Rejang, the largest river in Sarawak, and spread along the coast from the mouth of the Rejang northeast to the Tatau or thereabouts. The "core" dialects of Melanau form a chain, but from one end of the chain to the other there is a fairly high degree of divergence. Moreover, a number of other languages have at various times been identified as "dialects" of Melanau, and there continues to be considerable debate as to which groups do or do not belong in this category.

The "Orang Ulu" (a Malay phrase meaning 'up-river people') is the most diverse category, both linguistically and geographically. It includes literally hundreds of distinct dialect groups, most of them very small. However, the label "Orang Ulu" is purely political, and has no linguistic significance. The major ethno-linguistic groupings covered by this term include the KAYAN, the KENYAH, and the nomadic PENAN and PUNAN groups, all located in the Upper Rejang and Upper Baram basins, as well as

in Central Kalimantan; the “KAJANG”, in the Upper Rejang; the languages of the **LOWER BARAM**; the **KELABITIC** (or **APO DUAT**) languages, including Lun Bawang and Kelabit, at the extreme eastern end of the state; and the **BISAYA**, located in northeastern Sarawak, Brunei, and southwestern Sabah.

Of the works mentioned above, Blust’s study is especially helpful because he presents not only his conclusions but also a large part of the data on which those conclusions are based. In particular, he presents a complete matrix of cognate percentages for 56 languages and dialects (4 from the Philippines, 2 from Sabah, 3 from southern Kalimantan, and the rest from Sarawak, covering roughly the northern half of the state), together with the actual wordlists on which the calculations were based.

A rough indication of the degree of difference among the state’s language groups is given by the table of cognate percentages, extracted from Blust (1974a), shown in (2). Two words are considered to be **COGNATE** if they are descended historically from the same original word in a common ancestor language. For example, Malay *beras* and Sabah Dusun *wagas* are cognate forms, because they are both derived from the same ancestor form through regular sound changes. The higher the percentage of cognate forms, the closer the degree of similarity between a given pair of languages. The maximum possible value, of course, is 100%.

The percentages in figure (2) were calculated for a 100-word subset of the standard 200-word Swadesh list. “PSC” stands for “percentage of shared cognates”.

(2) **PSC figures from Blust (1974a)**
(rounded off to the nearest percentage point)

Selako (Kg. Pueh) ³							
29	Melanau (Mukah)						
27	45	Bintulu					
23	37	41	Berawan (Long Terawan)				
22	33	39	42	Narum (Lower Baram)			
24	37	36	41	32	Kenyah (Long Dunin)		
20	35	32	35	29	41	Kayan (Uma Juman)	
19	29	28	34	32	40	33	Kelabit (Bario)

What do these numbers tell us? First, note the relatively high PSC between Kenyah and Kayan (41%); the possible relationship between these two groups will be discussed in section 3.1. Second, we can see a slight elevation of scores among Melanau, Bintulu, Berawan and the Lower Baram languages (represented here by Narum). Whether this cluster represents a true linguistic subgroup is an interesting question, which will be discussed in section 3.2. It should be pointed out that the Kenyah scores seem slightly higher than expected with almost every other language,⁴ indicating perhaps some history of borrowing due to contact (direct or indirect) between Kenyah and the other groups. Allowing for this elevation of the Kenyah scores, we can see that the Kelabit group appears to be quite distinct from everything else in the state.⁵

Selako, as noted in note 1, is a Malayic Dayak language; thus the similarity of the other languages to Iban or Malay (not included in Blust's study) should be at about the same level as their similarity to Selako. Blust also did not include any Land Dayak wordlist in his cognate comparisons. I have attempted to estimate how the Land Dayak languages compare with the others listed in (2) by comparing data from Biatah Land Dayak with the wordlists included in Blust's study. This comparison was only partly successful, due to uncertainties about cognate relationships, synonyms, apparent borrowings, etc.; but the range of PSC values in the following table are enough to give some idea of the relative degree of lexical similarity.

(3) **PSC data from Blust (1974a)**
(Biatah data added from Topping, Kroeger)

Biatah (Kg. Kuab)								
28	Selako (Kg. Pueh)							
31-34	29	Melanau (Mukah)						
26-30	27	45	Bintulu					
26-27	23	37	41	Berawan (Long Terawan)				
22-25	22	33	39	42	Narum (Lower Baram)			
29-30	24	37	36	41	32	Kenyah (Long Dunin)		
26-28	20	35	32	35	29	41	Kayan (Uma Juman)	
21-27	19	29	28	34	32	40	33	Kelabit (Bario)

These figures suggest that Land Dayak, like Kelabit, is also quite distinct from all of the other language families in Sarawak, but slightly closer to Melanau than to the rest.