The Making of the Kadazan Dictionary: A Project Report

Rita Lasimbang

Project Manager - Kadazan Dictionary Project [1]
Department of Sabah Museum and State Archives
Sabah, Malaysia

1. The Kadazan Language

Kadazan (also known as Dusun) is the largest language community in the state of Sabah, Malaysia. Kadazan is a member of the Dusunic language family, North Bornean stock, Western Austronesian superstock (See Appendix for Map: Languages of Sabah, SIL 1984).

The Kadazan/Dusun language has 13 notable dialects with more than 300,000 speakers living in parts of 13 of the 23 administrative districts of Sabah. There are heavy concentrations of Kadazan/Dusun in the districts of Ranau, Tambunan, Penampang, Papar, Tuaran, Kota Belud, and those parts of Kota Kinabalu outside the city. Speakers of the Kadazan/Dusun language also live in Beaufort, Kinabatangan, Labuk-Sugut, and Keningau districts, with some migrant villages in the districts of Tenom and Tawau (Banker and Banker, 1984).

2. Orthography

The writing system which has been commonly used for the Kadazan/Dusun language follows quite closely the sound contrasts in the language. This one-to-one correlation between symbol and sound has made it easy for Kadazan speakers to learn to read and write their language, as well as for non-Kadazan language learners to reproduce, with reasonable accuracy, material written in Kadazan.

There are two exceptions to this general rule in the Coastal dialect. The glottal stop has generally been written only when it occurs in the middle of words. Only recently it has begun to be written when it occurs at the end of words. Also, the non-implosive **b** and **d** are generally not distinguished from the more common implosive **b** and **d**. Both these distinctions are important to proper understanding and speaking of the language. In order to make the dictionary as helpful as possible to students of the language, these distinctions have been marked in this dictionary. Glottal stop has been symbolised by (') and the implosive **b** and **d** by **b̲** and **d̲** respectively.

Kadazan has the following consonants and vowels:

b, b̲, d, d̲, g, h, j (loan words), k, l, m, n, ng, p, s, t, v, z, ' (glottal stop)

a, i, o, u

1 The author wishes to acknowledge the help in this project of Richard Brewis (SIL), Maria Benedict Lasimbang (KDCA) and Doris Lasimbang (KDCA).

1994. Bonifacio Sibayan & Leonard E. Newell, eds. Papers from the First Asia International Lexicography Conference, Manila, Philippines, 1992. LSP Special Monograph Issue, 35. Manila: Linguistic Society of the Philippines. pp. 53-57

3. History of the Project

Discussions about work on a Kadazan dictionary have been in process for quite some time. As stocks from the earlier *Kadazan Dictionary and Grammar* by the Rev. A. Antonissen (1958) were depleted, desire was expressed for an updated and expanded dictionary of the language. The reason for wanting an update is that in the Antonissen dictionary, words are listed in alphabetical order with English glosses, whereas this KDCA dictionary is a root oriented dictionary.

Acting on requests from the local people, the Kadazan Dusun Cultural Association (KDCA) in 1987 requested the help of the Summer Institute of Linguistics (SIL) researchers John and Carolyn Miller to serve as consultants for a Kadazan dictionary project. Available at the start of the project was the Kadazan section of the Antonissen dictionary which has been keyboarded onto computer by a local volunteer working with SIL, as well as the root form of each word as analyzed by John and Carolyn Miller when they were studying the Kadazan language.

4. The Database

The information which forms the database for the dictionary project has been entered onto computer with standard format markers, i.e., every type of information has been coded. The codes used are as follows:

\lx marks the Kadazan root word and the beginning of a main entry.
\et the etymology of those words known to be borrowed from another language—e.g. English or Malay.
\n the meaning number where a root has more than one unrelated meaning.
\id an idiom using the root word.
\co a comment expanding on the meaning.
\ms an empty field serving only to mark each expanded form.
\eg an expanded form of the root word.
\g1 an English gloss giving the meaning either of the root word or the expanded form.
\g2 a Malay gloss giving the meaning either of the root word or the expanded form.
\g3 the Central Dialect equivalent form of the root word.
\e# a number designating the combination of affixes which occurs in a particular expanded form.
\oa the breakdown of the expanded form into its various parts.
\og the grammatical designations for the affixes of the expanded form.
\ex a Kadazan sentence illustrating the usage of either the root word or the expanded form.
\tr an English translation of the Kadazan illustrative sentence.
\t2 a Malay translation of the illustrative sentence.

Size of database:

Number of files (A - Z)	=	132
Number of root words (\lx)	=	3,819
Number of expanded forms (\eg)	=	27,314

1994. Bonifacio Sibayan & Leonard E. Newell, eds. Papers from the First Asia International Lexicography Conference, Manila, Philippines, 1992. LSP Special Monograph Issue, 35. Manila: Linguistic Society of the Philippines. pp. 53-57

5. Description of Dictionary entries

Because of the complex affixation of words in languages such as Kadazan, words are entered by the root. In some cases, the root word never occurs by itself without affixation. Where this is the case the root has been marked by an asterisk (*).

The root word and the affixed forms of the word are given in the Coastal dialect. However, included with the root word is a Central dialect form of the root word (marked by C) as well as the gloss of the root word in Malay (marked by M) and English (marked by E).

With each root word are included the most common affixed forms of the root.

6. The Glossary

Kadazan has vowel-harmony rules and other morphophonemic changes that cause difficulty in identifying the root form of a word. A glossary will therefore be included in the back of the volume, giving a listing of every word in the dictionary with a breakdown of the form into its various parts including the root. Following the breakdown is given the function of each of these parts. Further information about the form and other forms of the same root may be found by looking up the root form in the main section of the dictionary.

Additional information about the Kadazan language may be found in the introduction to the glossary. This includes a description of regular sound changes, combinations of affixes and their usage, and a brief description of the use of focus.

7. Conclusion

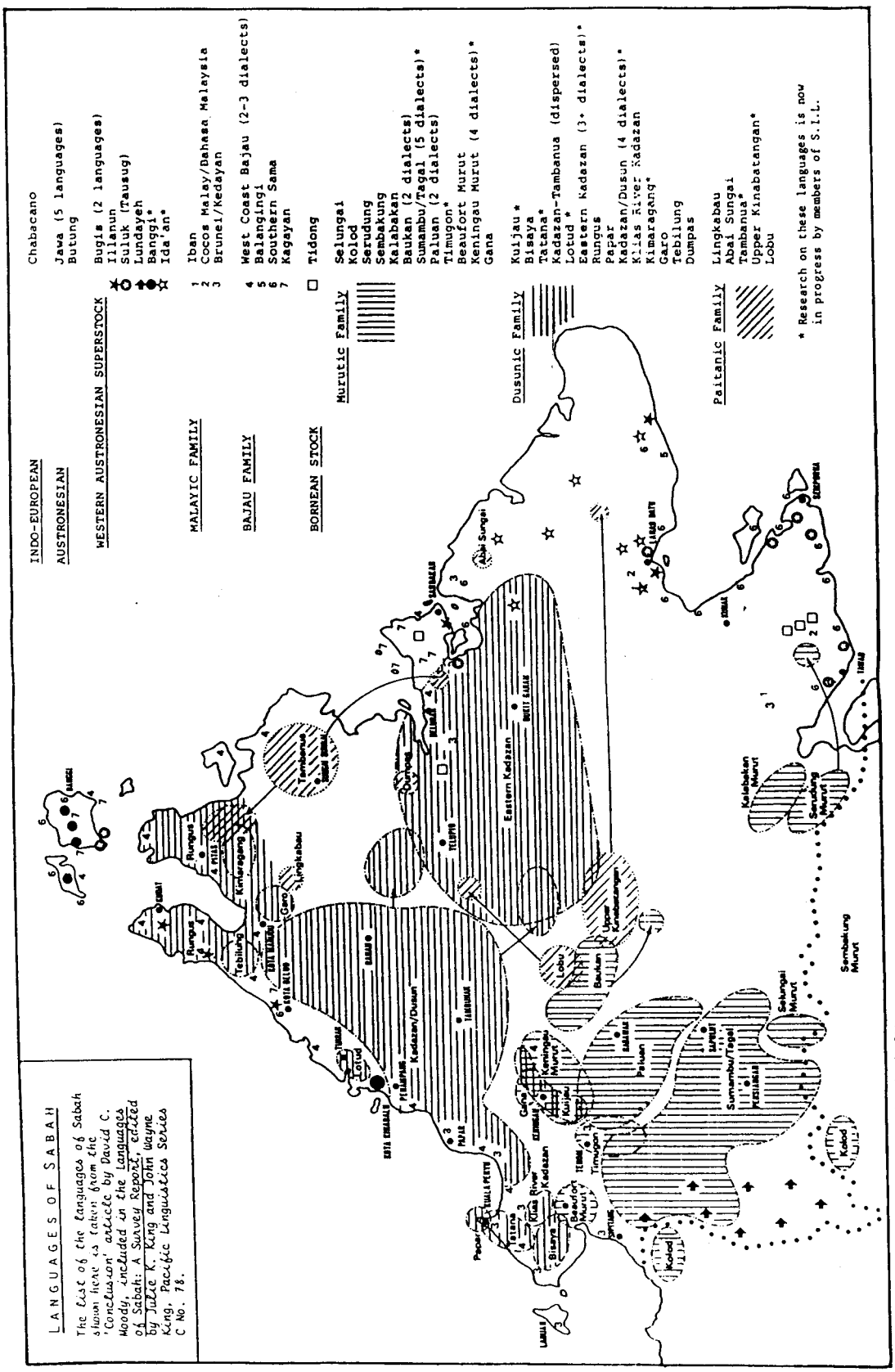
This dictionary is being prepared and designed primarily for use by the Kadazan community as an important step in the development the Kadazan language.

As it also contains important technical and grammatical information, English and Malay speakers will find the dictionary to be an aid in understanding the Kadazan language, and linguists interested in the languages of the area will find it a source of information for language comparison.

While members of the dictionary staff know that no dictionary is ever complete or without error, it is their hope that future generations of Kadazan speakers will add to it, correct it, and use the rich database for further language development.

1994. Bonifacio Sibayan & Leonard E. Newell, eds. Papers from the First Asia International Lexicography Conference, Manila, Philippines, 1992. LSP Special Monograph Issue, 35. Manila: Linguistic Society of the Philippines. pp. 53-57

APPENDIX



1994. Bonifacio Sibayan & Leonard E. Newell, eds. Papers from the First Asia International Lexicography Conference, Manila, Philippines, 1992. LSP Special Monograph Issue, 35. Manila: Linguistic Society of the Philippines. pp. 53-57

REFERENCES

Antonissen, The Rev. A. 1958. *Kadazan Dictionary and Grammar.* Canberra: Government Printing Office.

Banker, John and Elizabeth. 1984. "The Kadazan/Dusun language." Edited by J.W. King and J.K. King. Languages of Sabah: a survey report. *Pacific Linguistics C-78*: 297 -324 Canberra: The Australian National University.

Miller, John and Carolyn. 1988. Report to the Kadazan Dictionary Committee. Unpublished manuscript.

1994. Bonifacio Sibayan & Leonard E. Newell, eds. Papers from the First Asia International Lexicography Conference, Manila, Philippines, 1992. LSP Special Monograph Issue, 35. Manila: Linguistic Society of the Philippines. pp. 53-57