A Katuic cultural reconstruction

Anthony JUKES & Ilia PEJROS
Department of Linguistics,
The University of Melbourne
Parkville, VIC 3052 Australia

By analysing a proto-language it is possible for linguists to compile a so-called 'linguistic account' of the life of its speakers. Such an account based solely on linguistic data is by no means a complete description of prehistoric life, but only its reflection in the language. To obtain a comprehensive picture we need to bring together archaeological, linguistic, anthropological and other accounts, in order to identify their similarities and explain their discrepancies. In this paper we discuss part of the linguistic account of Katuic prehistory. This means that we are dealing here only with linguistic reconstructions and thus do not attempt to compare them with any extralinguistic facts.

A linguistic account of prehistory is based on extensive comparative research in the history of the chosen language family. The obligatory conditions required for completion of a linguistic account include:

- a detailed etymological dictionary of the family based on thorough comparison of each pair of languages studied;
- a precise comparative phonology of the family;
- a well-proven genetic classification of the family;
- a cultural-contact classification of the family which identifies cultural zones among related languages and thus predicts the possibility of borrowings within the family;
- a technique for identifying borrowings from unrelated languages.

The Katuic Etymological Dictionary (Pejros 1996) is based on a comparison of four main Katuic languages (Bru, Kui, Pakoh and Katu) with at least 1,000 roots known for each of them. The etymological dictionary was compiled through direct comparison of each pair of these languages (Bru and Kui, Bru and Pakoh, Bru and Katu, Kui and Pakoh, Kui and Katu, and Pakoh and
Katu) and is based on forms found in any two, three or in all the languages. Such a process allows us to believe that the majority of reliable etymologies has been discovered.

Pejros (1996) has established a set of systematic phonological correspondences between the four languages and has suggested a new Proto Katuic reconstruction which differs considerably from those previously proposed (Thomas 1967; Diffoth 1982; Efimov 1983).

The phonological correspondences do not provide any information that helps in classifying the languages, so, on the basis of phonological correspondences, we can treat the languages as four independent branches of the family. A lexicostatistical analysis\(^1\) gives the following matrix:

<table>
<thead>
<tr>
<th></th>
<th>Katu(^2)</th>
<th>Bru</th>
<th>Kui</th>
<th>Pakoh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katu</td>
<td>-</td>
<td>45%</td>
<td>39%</td>
<td>46%</td>
</tr>
<tr>
<td>Bru</td>
<td>45%</td>
<td>-</td>
<td>54%</td>
<td>56%</td>
</tr>
<tr>
<td>Kui</td>
<td>39%</td>
<td>54%</td>
<td>-</td>
<td>45%</td>
</tr>
<tr>
<td>Pakoh</td>
<td>46%</td>
<td>56%</td>
<td>45%</td>
<td>-</td>
</tr>
</tbody>
</table>

The best interpretation of the matrix is the following tree:

```
    Proto Katuic
      /    \
     /      \
   Katu    Pakoh
         /     \
        /       \
       Bru      Kui
```

The classification shown in this tree is based on the assumption that there has been borrowing between Bru and Pakoh since the break up of their mother language while Kui speakers have had little contact with speakers of their sister languages.

---

1 It is important to mention that the procedure of this lexicostatistical analysis is rather different from the standard one (see Pejros, to appear).
2 Note that figures given for the Katu language in this table are for the An Diem dialect.
This lexicostatistical classification correlates well with a lexical classification, which separates Katu from the other three languages. Cultural-contact considerations bring together the pairs Bru / Kui and Pakoh / Katu, suggesting that they belong to two different contact zones.

The Katuic languages were, and still are subject to intensive influences from Thai, Vietnamese, and perhaps also Khmer. Some borrowings came into Proto Katuic, but the majority have been adopted independently by the daughter languages.

The Katuic languages form a branch of the larger Mon-Khmer family, which in turn can be regarded as a branch of the Austroasiatic language family. Another branch of Austroasiatic is formed by Munda languages. This binary split is not, however, supported by sufficient comparative evidence. No reliable information is available about Proto Mon-Khmer: there is no detailed reconstruction of its phonology or morphology and its lexicon remains mostly unknown. As a result the genetic classification of the family remains obscure. It is possible, however, that Katuic languages are closer to Khmer and Vietic than for example Monic or Palaungic.

Given our present understanding of Katuic linguistic history it is often impossible to determine at what stage a word has been borrowed. Even if there is a similar word in Vietnamese or Khmer we cannot prove that it has been borrowed: the word could simply be of Proto Mon-Khmer origin and have been retained in the Katuic languages. At this stage of investigation the Katuic etymological dictionary includes both words of common origin which can be traced back to Proto Mon-Khmer, and ancient borrowings into Proto Katuic. To separate these two types of comparisons we need a detailed Mon-Khmer phonological reconstruction, which is not yet available.

One of the major sources of information for a linguistic account of prehistory is the lexicon of the proto-language. In it one can find words whose meanings are related to:

- various features of environment, like names of plants, animals, natural phenomena, etc.;
• various aspects of material culture;
• various aspects of social organisation and traditional beliefs.

In this paper we discuss only the information on the material culture which can be found in the reconstructed Proto Katuic lexicon. It is worthwhile to mention, however, that the dictionary is practically without the 'environmental' lexica; this phenomenon requires additional discussion (Pejros 1994).

The quality of data used in cultural reconstructions depends on two factors:
(i) the reliability of reconstructed meanings. With a meaning reconstructed too broad or too narrow our conclusions can be jeopardised. If, for example, we have a word which means 'duck' it is very important to know whether its real meaning is 'domesticated duck' or 'wild duck', without this information a precise semantic reconstruction cannot be made.

(ii) the reliability of stratification of reconstructions. In each case we need to know whether a form can be attributed to the Proto language level or to more recent levels of the family's classification. In theory two types of words can be included in a proto-language lexicon:
– words whose reflexes are found in main branches of the family. In the Katuic case we can be sure that a form belongs to the proto-language if it is represented in all four languages or in Katu and any other two and it is not a loan;
– words represented in one or two languages of the family, but also found in languages and branches related to the family under investigation. If a word is found (for example) only in Katu, but it is also known in other Mon-Khmer branches, we can attribute it to the Proto-Katuic level provided that it was not a loan in Katu.

As Proto-Mon-Khmer comparative phonology and lexicon are not yet known in detail, in this paper we will limit ourselves only with the forms whose proto-language stratification is supported primary by inner Katuic sources. External Mon-Khmer evidence will be added at the next stage of the investigation.

Cultural reconstructions are based on two interrelated postulates: