On the Centrality of Katuic-Bahnaric to Austroasiatic

What is Katuic-Bahnaric, and why does it have all those ties to the rest of Austroasiatic?*

Franklin E. Huffman
Cornell University

In an article in 1976 entitled 'The relevance of lexicostatistics to Mon-Khmer languages' (Huffman 1976), I attempted to show that the historical and universal assumptions of classical Swadesh glottochronology (Swadesh 1952) are invalid, as least as far as Austroasiatic languages are concerned, but that lexicostatistics as defined by Hymes (Hymes 1960), i.e. the simple quantification of cognates sharing a common gloss, is useful in showing relative distance between languages within a given family, provided one uses a constant corpus of vocabulary basic to that particular family. With all these reservations about lexicostatistics, I present this paper not as a serious or publishable piece of research, but rather because of its very surprising and intriguing results, and because, given the fledgling state of comparative Austroasiatic studies, it may be interesting faute de mieux.

By way of background, the Austroasiatic language family, first established by Schmidt in 1906, comprises some 150 languages spoken by minority groups in an area stretching from Eastern India through Burma, Thailand, Laos, Cambodia, Vietnam, and the Malay Peninsula, and includes the Khmer and Vietnamese national languages. In 1970, Thomas and Headley, on the basis of lexicostatistics, proposed nine subgroups for the Mon-Khmer branch: Monic, Khmeric, Pearic, Bahnaric, Katuic, Khmuic, Palaungic, Khasi, and Viet-Muong, leaving Munda and Nicobarese of India and the Malaccan languages of the Malay Peninsula as separate branches of Austroasiatic. In 1976 (Diffloth 1976b) he proposed the further addition of Nicobarese and Mang (which Perles 1974 separates from Viet-Muong) to Mon-Khmer proper, leaving Munda as the only branch outside Mon-Khmer, but in another 1976 article (Diffloth 1976a) he states that it is not clear whether the three Aslian groups are more closely related to each other or individually to other Mon-Khmer groups.

Now, it is entirely possible that there is a major cleavage in Austroasiatic between Munda and everything else, but it seemed to me that if everything else belonged in Mon-Khmer, maybe we could get Munda in too, discard the term Mon-Khmer (which is purely arbitrary anyhow), and talk only about branches of Austroasiatic until a more accurate subclassification can be established. So last year, as a project for my seminar in Austroasiatic linguistics, and to satisfy my personal curiosity as to just how close Munda, Nicobarese, and Aslian are to the traditional Mon-Khmer languages, I decided to extend my previous lexicostatistical comparison to all branches of Austroasiatic.

In my 1976 article I had devised, in addition to a 500-word list, a 100-word list of really hard-core basic vocabulary; I decided to use this same 100-word list for consistency of comparison, and because it was impossible to find equivalents for the 500-word list in the meager literature on the more exotic Austroasiatic languages. From the eight branches of Austroasiatic I had covered in 1976, I used Mon, Khmer, Lawa representing Palaungic, Thin representing Khmuic, Kuy representing Katuic, Loven and Stieng from Bahnaric, since the subgrouping of Stieng has been somewhat controversial (i.e. considered Khmeric by Shorto, Jacob, and Simmonds 1963, but Bahnaric by Thomas and Headley 1970), Chong

*Working paper; not for publication.
representing Pearic, and Vietnamese; to these I added Muong (from Barker and Barker 1972 and Huffman 1977), Jehai, Semai, and Semelai from Northern, Central, and Southern Aslian respectively (Benjamin 1976 and Diffloth, personal communication), Khasi (from Amirkha 1931 and Rabel 1961), three Munda languages - Remo (from Bhattachariya 1968), Kharia (from Floor, Gheysens, and Druart 1934 and Pinnow 1959), and Ho (from Deeney 1975). - and finally Nicobarese (from Man 1889 and Whitehead 1925).

Thus I had a corpus of 100 words of basic vocabulary for each of 18 languages representing 14 branches of Austroasiatic. I then proceeded to figure the interlanguage cognate percentages for the 153 possible pairs of 18 languages; this comparison produced some interesting results:

1) It became obvious that the cognate percentages between the Katuic and Bahnaric branches were sufficiently high (52-55) and sufficiently similar to be considered a single branch at the interbranch level of comparison; in fact the cognate percentage between Loven of Bahnaric and Kuy of Katuic was one percentage point higher than that between Loven and Stieng, the other Bahnaric language; these were furthermore the highest percentages in the entire chart, except between Vietnamese and Muong, which showed a cognate percentage of 89.

2) I had been chastised in my previous study for not including Muong in my calculations, which, it was claimed, would have put Viet-Muong more squarely within Mon-Khmer, and in fact the Muong percentages averaged three percentage points higher than those for Vietnamese.

3) Semelai of Southern Aslian showed a higher cognate percentage with the Katuic-Bahnaric languages than with Jehai, its Northern Aslian congener, which reflects Diffloth's indication as to whether there should be three separate branches or a single branch of Aslian; however it is significant that Semelai's closest relationship is with the Katuic-Bahnaric languages, as I will show later.

4) Within Munda, the cognate percentages for Kharia averaged six percentage points higher than those for Remo and Ho; this reflects, I fear, the higher degree of sophistication of the source

used for Kharia, namely Pinnow's Kharia-Sprache (Pinnow 1959), in which comparative reconstructions are suggested for Kharia cognates with other Austroasiatic languages. This illustrates one of the pitfalls of lexicostatistical comparison; the cognate percentage obtained will vary in proportion to the investigator's knowledge of the phonological changes which have taken place in the languages involved.

In order to level out the skewing effect of the higher Kharia figures in Munda, and of the higher interlanguage figures in Aslian, Viet-Muong, and Katuic-Bahnaric, I decided to average the cognate percentages obtained in those branches represented by more than one language, resulting in eleven branches. I then figured the 55 interbranch cognate percentages resulting from eleven branches, and further calculated the average cognate percentages for each branch. (See Chart 1). This operation produced two very surprising and interesting results:

1) The eleven branches of Austroasiatic form a continuum of relatedness within Austroasiatic as a whole, from Katuic-Bahnaric through Nicobarese. Now average cognate percentages would be irrelevant if they were the result of widely divergent figures, but the significant fact is that the Katuic-Bahnaric percentages are, with two exceptions, consistently the highest, while the Nicobarese percentages, also with two exceptions, are consistently the lowest, with the branches in between showing a similar curve. (See Chart 1).

2) But the more intriguing result of these calculations is the fact that eight of the ten branches of Austroasiatic compared with Katuic-Bahnaric showed their highest cognate percentage with Katuic-Bahnaric, and the remaining two showed their second highest cognate percentage with Katuic-Bahnaric: In other words, except for Pearic, which is closest to Khmer and next closest to Katuic-Bahnaric, and Nicobarese, which is equally close to Khmer and Palaungic and next closest to Katuic-Bahnaric, Katuic-Bahnaric appears to be more closely related to all branches of Austroasiatic than are those branches to each other, regardless of geographic location. (See Chart 2). Now what one would logically
expect, in the abstract, is a chain effect whereby Munda, for example, would be most closely related to one of its closest neighbors, such as Khasi or Nicobarese, then to Palaungic, then to Mon, and so on in an easterly direction; what we in fact have, on the contrary, is that Munda is most closely related to Katuic-Bahnaric, next to Khmer, next to Mon, and so on in a westerly direction, and shows its lowest cognate percentages with its closest geographical neighbors, Khasi and Nicobarese: This would seem to argue for an eastern (Central Vietnam) center of dispersal and a separate westward migration for each branch of Austroasiatic.

Now I have no intention of publishing these findings or even of defending them very strongly, for the following reasons:

- the corpus used is too small to be very reliable;
- lexicostatistical judgements are highly subjective, and differ widely depending on how much one knows about the phonological history of the languages involved; and
- subgrouping must ultimately be based on comparative reconstruction of individual branches.

I present these results to this conference merely for the consideration, evaluation, and possible amusement of my colleagues.

Bibliography


Diffloth, Gérard, 1976a. 'An appraisal of Benedict's views on Austroasiatic and Austro-Thai relations', Discussion paper no. 82, Center for Southeast Asian Studies, Kyoto Univ., March 1976.

Diffloth, Gérard, 1976b. 'Proto-Mon-Khmer final spirants', Discussion paper no. 86, Center for Southeast Asian Studies, Kyoto University, June, 1976.


Hymes, D. H., 1960, 'Lexicostatistics so far', *Current Anthropology* 1:2:34-44.


Handout

On the Centrality of Katuic-Bahnaric to Austroasiatic

Huffman

Chart 1: Interbranch Cognate Percentages and Averages

<table>
<thead>
<tr>
<th></th>
<th>K-B</th>
<th>Kme</th>
<th>Mon</th>
<th>Pal</th>
<th>V-M</th>
<th>Pea</th>
<th>Asl</th>
<th>Kmu</th>
<th>Kha</th>
<th>Mun</th>
<th>Nic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katuic-Bahnaric</td>
<td>47</td>
<td>42</td>
<td>34</td>
<td>33</td>
<td>35</td>
<td>31</td>
<td>32</td>
<td>24</td>
<td>27</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Khmer</td>
<td>47</td>
<td>33</td>
<td>31</td>
<td>30</td>
<td>39</td>
<td>28</td>
<td>25</td>
<td>24</td>
<td>24</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Mon</td>
<td>42</td>
<td>33</td>
<td>33</td>
<td>30</td>
<td>22</td>
<td>24</td>
<td>28</td>
<td>22</td>
<td>20</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Palaungic</td>
<td>34</td>
<td>31</td>
<td>33</td>
<td>26</td>
<td>25</td>
<td>22</td>
<td>26</td>
<td>26</td>
<td>18</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Viet-Muong</td>
<td>33</td>
<td>30</td>
<td>30</td>
<td>26</td>
<td>24</td>
<td>22</td>
<td>24</td>
<td>23</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Pearic</td>
<td>35</td>
<td>39</td>
<td>22</td>
<td>25</td>
<td>24</td>
<td>24</td>
<td>22</td>
<td>17</td>
<td>19</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Aslian</td>
<td>30</td>
<td>28</td>
<td>24</td>
<td>22</td>
<td>22</td>
<td>24</td>
<td>24</td>
<td>20</td>
<td>17</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Khmuic</td>
<td>32</td>
<td>25</td>
<td>28</td>
<td>26</td>
<td>24</td>
<td>22</td>
<td>24</td>
<td>24</td>
<td>20</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Khasi</td>
<td>24</td>
<td>24</td>
<td>22</td>
<td>26</td>
<td>23</td>
<td>17</td>
<td>20</td>
<td>24</td>
<td>18</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Munda</td>
<td>27</td>
<td>24</td>
<td>20</td>
<td>18</td>
<td>20</td>
<td>19</td>
<td>17</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Nicobarese</td>
<td>20</td>
<td>23</td>
<td>19</td>
<td>23</td>
<td>18</td>
<td>16</td>
<td>19</td>
<td>17</td>
<td>14</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Interbranch totals</td>
<td>324</td>
<td>304</td>
<td>273</td>
<td>264</td>
<td>250</td>
<td>243</td>
<td>231</td>
<td>222</td>
<td>212</td>
<td>200</td>
<td>186</td>
</tr>
<tr>
<td>Interbranch averages</td>
<td>32.4</td>
<td>30.4</td>
<td>27.3</td>
<td>26.4</td>
<td>25.0</td>
<td>24.3</td>
<td>23.1</td>
<td>22.2</td>
<td>21.2</td>
<td>20.0</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Chart 2: Geographical Location and Closest Affiliation