Some Evidence Favoring the Central Hypothesis

Isidore Dyen
Yale University (Emeritus)

The Central hypothesis proposes that the Austronesian homeland was in the neighborhood of New Guinea. It combines the first two alternatives I proposed in 1965 (1965.57), Melanesia-East New Guinea and West New Guinea, as against the third alternative, Taiwan, the locale of the Formosan languages. These hypotheses were reached by lexicostatistics. The third alternative is now called the Formosan hypothesis and is supported by Blust and, we are frequently told, a majority of Austronesianists.

1. The homomeric method and Blust’s charge of circularity

My use of the so-called homomeric method has been criticized by Blust (1999.67) as based on circular reasoning. The homomeric method draws inferences leading to subgroups from collections of cognate sets with the same distribution over languages. Such a collection is called a homomery from ‘homo-’ = same, ‘-mer-’ = measure. A homomery is a collection of exclusively shared cognate sets.

Blust’s formulation of what he presents as ‘Dyen’s argument’ (ibid.) lacks a citation from my publications because I never published such a statement. The critical part is completely contrived, and is the worse because it can be interpreted as a quotation or paraphrase of an argument that I made concerning the use of homomeries to infer a subgroup. Since he does not cite a source, and I do not recognize it as representing my thinking, it can only be regarded as his own concoction that some may consider unethical. His formulation follows verbatim:

‘...Dyen’s arguments (no citation - ID) for a Formosan subgroup:
‘A = Formosan languages form a subgroup, B = they share some ‘homomeries’, C = they have undergone a period of exclusively shared history.
‘It follows that Formosan languages form a subgroup because they share some homomeries; they share some homomeries because they have undergone a period of exclusively shared history; and they have undergone a period of exclusively shared history because they form a subgroup. In short, Formosan languages form a subgroup because Formosan languages form a subgroup.’

It is not only fallacious, but also falsely attributed. The invalidity and contrivance in his formulation can be seen by applying his model to his later suggestion that Formosan languages do not form a subgroup because many do not share phonemic mergers as follows:

‘A = Formosan languages do not form a subgroup, B = many do not share significant phonemic mergers, C = they have not undergone a period of exclusively shared history.
‘It follows that Formosan languages do not form a subgroup because many do not share significant phonemic mergers; they do not share phonemic mergers because they have not undergone a period of exclusively shared history; and they have not undergone a period of exclusively shared history because they do not form a subgroup. In short, Formosan languages do not form a subgroup because Formosan languages do not form a subgroup.’

My argument for a Formosan subgroup on the basis of homomeries is the following and it is not circular:
The Formosan languages form a subgroup if and only if they continue the same mesolanguage, a period of exclusively shared history. The likelihood that a mesolanguage existed is commensurate with the weight of the evidence of traits exclusively shared by Formosan languages, including homomeries, that imply innovations in that mesolanguage.

2. The Formosan hypothesis and shared phonemic mergers

Blust’s newest argument in support of the Formosan hypothesis is based on the observation that many Formosan languages share no significant phonemic mergers (1999.37-55). He couples this with the proposal that only shared phonemic mergers should be used in subgrouping. Phonemic mergers have the advantage that they can not be reversed.

Shared phonemic mergers are highly regarded as evidence for subgrouping. Nevertheless their occurrence can not be guaranteed. The non-occurrence among immediate daughters of a proto-language can not be distinguished from the non-occurrence among the daughters of an immediate daughter. If many Formosan languages exhibit no significant shared mergers, some of them might yet be granddaughters that lacked shared mergers. If Proto-Austronesian could dissolve into daughters that showed no significant shared mergers, why should not a daughter dissolve in the same way?

A second objection arises from experience with the Indo-European languages. Proto-Iranian shares with Proto-Balto-Slavic the merger of voiced aspirate stops with plain voiced stops, but is nevertheless associated in Proto-Indo Iranian with Proto-Indic, which does not exhibit this merger. On the other hand the so-called centum languages are believed to have merged velar stops with palatal stops, but are not taken to form a subgroup. Similarly the satem-languages merge velar and labiovelar stops without being formed into a subgroup. In both cases the changes are interpreted as constituting isoglosses in Proto-Indo-European.

3. The likelihood of subgrouping among the Formosan languages

The following is essentially the same subgrouping Tsuchida and I presented at VICAL 2 (1991.92), differing only in grouping Kavalan with Ami:

Proto-Formosan.
1. Proto-North-Formosan.
  1.1. Proto-Atayalic.
    1.1.1. Atayal.
    1.1.2. Sediq.
  1.2. Saisiyat.
  1.3. Pazeh.
2. Proto-South-Formosan.
  2.1. Bunun.
  2.2. Thao.
  2.3. Proto-Puyumo-Rukaic.
    2.3.1. Proto-Rukaitsouic.
    2.3.1.1. Proto-Tsouic.
    2.3.1.1.1. Tsou
    2.3.1.1.2. Proto-South-Tsouic.
    2.3.1.1.2.1. Kanakanabu.
    2.3.1.1.2.2. Saaroa.
2.3.1.2. Proto-Rukaic.
2.3.1.2.1. Lower-three Rukai.
2.3.1.2.2. Rukai Proper.
2.3.2. Proto-Paiwanic.
2.3.2.1. Proto-Puyumo-Amic.
2.3.2.1.1. Kavalan.
2.4.2.1.2. Ami.
2.4.2.1.3. Proto-Puyumic.
2.4.2.1.3.1. Puyuma.
2.4.2.1.3.2. Paiwan.

Li (1990) offers two slightly different families, both of which group Kavalan with Ami and I agree with Li’s conclusion. I place Kavalan as coordinate with Ami and Puyumic in accordance with the evidence of the following lowest-level homomorphic grid that I have drawn up:

<table>
<thead>
<tr>
<th></th>
<th>Pai</th>
<th>Puy</th>
<th>Ami</th>
<th>Kav</th>
<th>Tsc</th>
<th>Rkc</th>
<th>Bun</th>
<th>Tha</th>
<th>Atc</th>
<th>Sai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puy</td>
<td>117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ami</td>
<td>46</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kav</td>
<td>12</td>
<td>3</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsc</td>
<td>18</td>
<td>2</td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rkc</td>
<td>32</td>
<td>11</td>
<td>6</td>
<td>2</td>
<td>166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bun</td>
<td>5</td>
<td>5</td>
<td>24</td>
<td>1</td>
<td>25</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tha</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>9</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atc</td>
<td>3</td>
<td>1</td>
<td>12</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sai</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Paz</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>17</td>
</tr>
</tbody>
</table>

This represents a first step in classification. Tsuchida (1976.9-10) concluded essentially that for the languages he dealt with there was a Proto-Formosan with two immediate subgroups: Atayalic and non-Atayalic. Later Li (1985) suggested that Saisiyat and Pazeh should be associated with Atayalic in a North Formosan and the remainder as South Formosan. On the basis of this division the homomorphic evidence for a Proto-Formosan would be approximately 265 cognate sets that contain at least one North Formosan member and one South Formosan member and no extra-Formosan member. This number does not include any sets with a cognate belonging to a defunct language, so that, if such sets do not affect the North-South division and are added, the final number might be larger.

4. **Proto-Philippine as the closest relative of Proto-Formosan**

I regard the Formosan languages as a single subgroup and the Philippine subgroup as their closest relative. My view was reached on finding about 475 cognate sets shared by the Philippine subgroup with Formosan languages. It stands in sharp contrast with the 25 cognate shared by Formosan languages only with Oceanic languages.

These numbers are not a list of innovations, but a list containing innovations, a subtle difference. The inclusion of a very large number of innovations is indicated by the magnitude of the collection. It was further indicated that removals from the list were to be expected as research continues, but that additions likewise are not unlikely. The subtraction of one or even a few
cognate sets by the finding of a nullifying cognate does not affect the value of a whole collection of sets.

My attention was recently drawn to 800 West Indonesian cognate sets privately shared with Philippine languages. This large number includes many that were culled from the writings of Blust. This number suggests a closer tie between the West Indonesian and the Philippine languages than between the latter and the Formosan languages. Though I treat ±800 as probative, its size may nevertheless be partly due to the large number of different languages in the Philippine and West Indonesian subgroups as compared with the Formosan.

The apparent contradiction between the classification and the reference to the ‘closest relative’ can be resolved under a hypothesis that the respective ancestors of the three subgroups formed a dialect chain with a Philippine ancestor dialect between the other two. The differentiation between the Formosan and West Indonesian began while they were separated from each other by the Philippine ancestor. Formosan separated from the Philippine ancestor and the separation of West Indonesian from Philippine occurred later. This hypothetical sequence of events would explain the fact that only 95 cognates have been found to be shared by West Indonesian languages only with Formosan languages. That is why it may be claimed that the Philippine languages are the closest relative of the Formosan languages.

However it is worth noting that, given the West Indonesian-Philippine unity implying a Proto-Indo-Philippine the West Indonesian-Formosan-only cognate sets should then be added to those pointing to a unity of the three subgroups, a Proto-Indo-Formosan. Furthermore now the 173 cognate sets shared only by Formosan languages with both West Indonesian and the Philippines should likewise be added to those pointing to a Proto-Indo-Formosan. The combined total of cognate sets supporting a Proto-Indo-Formosan has thus risen to 743.

5. Homomeric evidence for a Hesperonesian (West Austronesian)

It is difficult to believe that a dispersal from Taiwan took place before the introduction there of rice cultivation. If the date of dispersal has been placed as late as 3000 B.C.E., it clearly postdates by a few millennia the beginning of cultivation of rice, which has been dated as early as 7000 B.C.E. in East Asia. For this reason the fact that rice cultivation did not reach Oceania until modern times has been viewed as an obstacle to the Formosan hypothesis. The explanation offered for the absence of rice cultivation has been that the practice was abandoned in Oceania for some good reason, despite the difficulty of imagining what that might be.

It is striking nevertheless that a list of cognate sets associable with rice cultivation can be assembled each of which has a Formosan, a West Indonesian, and an East Indonesian member and no Oceanic member. The list includes: words for ‘rice plant’, ‘husked rice’, ‘a stage of rice’, ‘mortar’, ‘pestle’, ‘to pound’, ‘to thresh’, ‘to winnow’. (Cf. also Dyen 1992.189-191).

However these sets do not stand alone. All together over 100 such cognate sets have been found. Among them cognates for ‘dog’ are similarly distributed. The explanation for the non-appearance of a cognate in Oceania has been offered that the dog lost its economic value and was eaten to extinction. The explanation requires us to accept the proposition that food had no economic value.

Among the other cognate sets there are: 9 cognate sets for human body parts and actions, 13 for other animates, their parts and associated behaviors or actions, and 11 for plants, their parts and associated behaviors.
6. The source of the Melanesian divergence

The high degree of divergence among Melanesian languages has been generally recognized, there is some disagreement about its explanation. Ross reports (1988.23) that Lynch attributed the divergence to contact whereas Pawley attributed it to time-depth.

It is generally agreed that for most of Melanesia, the Austronesians were newcomers, proceeding from north to south, either as invaders or colonizers or both. They are distributed on islands and on coastal areas of Eastern New Guinea, extending inland only in a few places up rivers.

There is little linguistic that shows the direction of the migration. Even the delimitation of an Oceanic subgroup has offered difficulties. Certainly the lexicostatistical results do little more than define the problem. A solution to the problem set by the lexicostatistics would be that the Austronesian settlement took place so long ago that the evidence of the immigration has been so overlaid by population movements and linguistic change that the linguistic data do not reflect it. The contradiction between the evidence of the distribution of the languages, which points to an invasion, and the linguistic evidence, which points to a long occupation, can only be resolved by careful application of the comparative methods that are available in the expectation that the evidence will provide a basis for inferring a reasonable date of a Proto-Oceanic that differs from Proto-Austronesian.

The lexicostatistical cognate percentage of a pair of interrelated speech-types indexes the beginning of their divergence in their last protolanguage. The lower the shared percentage, the longer the pair has been diverging since their last protolanguage. The higher the shared percentage, the more recent the beginning of divergence was.

Swadesh inferred that his percentages were the product of a process analogous to radiocarbon decay (1952.453-454) and proceeded therefore to use the percentages as indicators of elapsed time. The likelihood of a replacement was not the same for each member of the basic vocabulary list. For example the likelihood is different for ‘two’ and ‘play’. The likelihood that a particular atom of radioactive carbon might lose its radioactivity is the same as that of any other atom. This difference was confirmed in Dyen, James, and Cole (1965).

This observation does not invalidate lexicostatistics nor necessarily using the rate of replacement as a crude indicator of elapsed time. The current forms when applied consistently to large numbers of languages, gives equal chances to the different entry-lists to exhibit degrees of relationship between their languages and these degrees reflect in a rough way different time intervals. Furthermore some inconsistencies in the results have been attributed with some success to the effects of borrowing and word-taboo (cf. Hymes 1960.8-9, Dyen 1967).

Lexicostatistical results with the current basic vocabulary lists are more informative when the percentages are above 20% rather than below it. The percentages below 20% can presumably be counted on to be made up in large part of vocabulary items that are more resistant to replacement than the rest, perhaps for example low numerals, personal pronouns, and close-kinship terms. At the same time it should be observed that demonstratives and conjunctions appear to undergo replacement at a much greater rate than the majority of entries. There is here the obvious suggestion that experiments are in order for the testing of basic vocabulary lists that call for entries with more nearly equal likelihoods of replacement.

The evidence of the high degree of divergence among Melanesian languages appears in the wide distribution of very low percentages. Percentages below 20% can be found in Tryon’s lexicostatistical comparison of New Hebrides languages with each other and in the later similar comparison of the Solomons languages.
Furthermore among the languages of the Morobe District of Papua-New Guinea Hooley reports many percentages at 10% or lower on the basis of a 100-item list and somewhat fewer on the basis of a 128-word list.

The members of some different Tip language subgroups share percentages below 20% as in the following comparison of six Tip languages:

<table>
<thead>
<tr>
<th></th>
<th>Motu</th>
<th>Sinaugoro</th>
<th>Kilavila</th>
<th>Muyuw</th>
<th>Molima</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinaugoro</td>
<td>40.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kilavila</td>
<td>12.3</td>
<td>12.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muyuw</td>
<td>12.4</td>
<td>13.1</td>
<td>48.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molima</td>
<td>13.8</td>
<td>13.1</td>
<td>16.5</td>
<td>17.6</td>
<td></td>
</tr>
<tr>
<td>Wedau</td>
<td>16.1</td>
<td>9.1</td>
<td>13.7</td>
<td>11.7</td>
<td>15.7</td>
</tr>
</tbody>
</table>

In New Caledonia I found on the basis of a 194-item list about 9% of cognate pairs between Canala and Nélémwa-Nixumwak; their respective percentages with Lenakel (on Tanna, southern New Hebrides) were found to be under 7%. Their percentages with Motu were slightly higher, but under 10%. On the other hand Motu scored about 14% with Lenakel.

However all these sets of languages are customarily assigned to the as yet ill-defined Proto-Oceanic. The widely offered hypothesis that contact could explain the extreme divergence among the languages of Melanesia is faced with two obstacles. Contact as it affects vocabulary in general and thus lexicostatistics appears as borrowing. There has been little or no progress in establishing the direct evidence of borrowing from the putative earlier inhabitants of Austronesian Oceania. Although it is true that there are a few instances of languages that can be shown to have borrowed extensively from non-Austronesian languages, these languages are by no means typical and simply bring to the fore the likelihood that among the languages that do not exhibit detectable borrowings there are many that have few undetectable borrowings, if any.

The second obstacle to the contact hypothesis is the nature of the borrowings that are to be expected. Since the migrating Austronesians survived, it is likely that they came in as the superior group in the contact that resulted. It can be surmised that their superiority was enforced by their use of the bow.

In such social contacts, borrowing is expected to be of the ‘intimate’ (Bloomfield 1933.461-475) or ‘prestige-seeking’ (Hockett 1958.404-405) variety. This variety of borrowing does tend to provide loanwords that end up in the basic vocabulary, but tends to be socially downward from, rather than upward into the language of the dominant group. Thus sufficient evidence of borrowing is lacking and the social conditions, as far as we can infer them to have been, do not favor the type of borrowing that would affect the basic vocabulary. The best hypothesis therefore appears to be that the low percentages found in Austronesian Oceania are more likely to reflect divergence associated with time of separation than borrowing associated with contact.

Ross subscribes to the hypothesis proffered by archeologists that Austronesians arrived in Central Papua about 100 B.C.E. (1988.195, 1994.391) and a millennium later underwent rapid cultural change. The collocation of rapid cultural change with rapid linguistic change seems hardly fortuitous, yet there is no necessary association between them. There is no clear evidence that rapid cultural change goes hand in hand with rapid linguistic change.

There is some archeological evidence that can be associated with early Austronesian penetration of Melanesia. Lapita pottery has often been associated by scholars with the advance of the Austronesians. It is therefore perhaps of some importance that Lapita pottery has been found on Ile-des-Pins off New Caledonia associated with a ‘radiocarbon date of 2190-2030 B.C.}
A date of 4000 B.P. at the very distant end of an Austronesian migration from Taiwan is hard to reconcile with a hypothesis that it began as late as 5000 B.P.

7. The double ‘express train’ consequence of the Formosan hypothesis.

Lexicostatistical results appear to fit with the proposition that a fair number of Formosan languages are distantly interrelated with each other and thus in all likelihood with any extra-Formosan language. Tsuchida’s diagram (1976.7) of lexicostatistical percentages shared by pairs of Formosan languages give a number of percentages below 20%. This observation appears to support the hypothesis that a fair number of Formosan languages were immediate daughters of Proto-Austronesian. However the two scholars that have been most deeply concerned with Formosan studies, Shigeru Tsuchida (1985) and Paul J. Li (1990), have published in favor of a Proto-Formosan.

The Formosan hypothesis could be the source of great gratification to Shigeru Tsuchida. His dissertation (1967) could then be seen to have more to do with the reconstruction of Proto-Austronesian than he thought. All cognate pairs shared by an immediate daughter of Proto-Austronesian with another Austronesian language, Formosan or not, is then automatically inferred to reflect a Proto-Austronesian etymon. If all or nearly all of such daughters were Formosan, his dissertation, which actually broached the study of the interrelations of a large number of Formosan languages, was a fundamental study in the reconstruction of Proto-Austronesian.

However low lexicostatistical percentages are not limited to the Formosan languages. The degree of divergence among some groups in Melanesia is very similar to that between the most divergent Formosan subgroups.

It might be expected that the percentages between a Tip language and a Formosan language would be much lower than their percentages with their respective neighbors. If that were true, it would follow that from the point of view of the Tip languages, the Formosan languages constitute a subgroup. Actually a Motu-Paiwan comparison yielded 16%, a percentage like Tsuchida’s 16% for the Atayal-Paiwan pairing.

The Formosan hypothesis regards Taiwan as the homeland, the site of Proto-Austronesian, and Proto-Oceanic as a subgroup within an extra-Formosan subgroup. To reach the site of Proto-Oceanic from Taiwan can reasonably be expected to take some time. After all, the expansion can hardly have been purposeful; the migrants could hardly have seen the end of the expansion as a goal. They were evidently used to the sea, but the dangers anticipated on leaving home would tend to make the expansion one that was a reaction to pressure than from a desire to pioneer. Yet that expansion had to reach the site of Proto-Oceanic in the east in time for the spread into the Tip of New Guinea so that the divergence of the Tip languages could begin; according to the lexicostatistics, the beginning of this divergence would already exist at approximately the same time that the Formosan languages began to diverge.

It has been generally realized that the findings in Austronesian Oceania under the Formosan hypothesis requires rapid language change immediately upon the arrival of the Austronesians in Oceania, by some described as ‘express train’ change. Actually under the Formosan hypothesis the interval from Proto-Austronesian to Proto-Oceanic according to Ross, Pawley and Osmond (ibid.) contains the succession: Proto Malayo-Polynesian, Proto Central/Eastern Malayo-Polynesian, Proto Eastern Malayo-Polynesian. If we set 700 years as an arbitrary minimum for the interval between successive proto-languages, the four stages would require at least (4x700=) 2800 years. If PAN is set at 3000 B.C.E., Proto-Oceanic would arrive at
Oceania close to the beginning of the first millennium C.E. This forces Oceanic languages to accomplish the equivalent of 3500 years of change in 2000 years. Even if this rapid change is equally divided between the Proto-Oceanic stage and later stages the result is 1750 years of change per thousand years, a rate of change that is twice what was assumed to be minimal.

What has not been recognized is that the so-called Oceanic languages require another ‘express train’ set of changes to account for their divergences among themselves. One way to avoid having the two ‘express trains’ proceeding in succession is to have them proceed simultaneously from a central source, say from somewhere in the neighborhood of New Guinea.

Actually the essence of the problem was recognized in Dyen 1965 where what is here called the Central hypothesis was presented as the first and thus the most likely explanation of the available data. It was however the recognition of the early date of the presence of the Formosan languages on Taiwan indicated by their percentages that led to the consideration of the findings by the homomeric method, in the hopes that it could provide decisive evidence.

It is difficult to measure the difference between the diversity among the Formosan languages and the diversity among the eastern languages if the lexicostatistics are ignored. Apparently either one could be the starting point of the Austronesian migration that is necessary to explain the Austronesian dispersal. The Central hypothesis has the advantage of requiring less time for the Austronesian diaspora. The Formosan hypothesis furthermore has the disadvantage of being closely tied to the Indo-Philippine subgroup by a very large homomery. Taken together they satisfy the critical requirement of a collection of closely related subgroups to be expected at the tail-end of a migration.
References


The preceding document was presented at the Tenth International Conference on Austronesian Linguistics (10-ICAL). To properly reference this work, please use the following format:


For other papers that were presented at 10-ICAL, please visit http://www.sil.org/asia/philippines/ical/papers.html.