Benedict's Austro-Tai Hypothesis—An Evaluation

LAWRENCE A. REID

INTRODUCTION

The possible connections among the hundreds of languages of Southeast Asia have been the subject of considerable research and a fair amount of speculation since at least the beginning of this century. Three major language families have generally been recognized—Austroasiatic (including languages such as Vietnamese, Mon, Kimer, Khasi, Nicobarese, and Munda), Sino-Tibetan (including Chinese, Tibetan, Burmese, and scores of other little-known languages), and Austronesian (represented in Mainland Southeast Asia by the Chamic group and Malay). Attempts to relate various of these families into supersets and to establish the position of Thai within one or another of these families dates back to Schmidt's (1906) attempt to relate Austronesian and Austroasiatic in a family that he called Austric. Thai, primarily because of its monosyllabic word structure and its tonal system which is similar to Chinese, was generally classified as belonging to a Sino-Thai group within Sino-Tibetan.

In 1942, Paul Benedict published a paper in the American Anthropologist that proposed what he called “a new alignment.” He proposed that Thai was not genetically related to Chinese at all, but to the Austronesian language family, which he did at that time, following Dempwolff (1938), called the Indonesian languages. A series of papers followed, culminating in a volume, Austro-Thai Language and Culture (Benedict 1975), which reprinted his earlier papers and included a glossary of over 200 pages of Austro-Thai reconstructions with extensive discussion of sound correspondences and lexical evidence from each of the major branches of his proposed family.

Despite a ringing endorsement from Ward Goodenough in the foreword, comparativists in the Austronesian field have been lukewarm at best towards the hypothesis. Two of the better-known linguists in this group, Isidore Dyen and Robert Blust, both reject it. Blust (personal communication) states, “the correspondences simply don’t work, unless you fudge every case with special conditions.”

The purpose of this paper is to take a close look at what Benedict has done and to provide an evaluation of his proposed hypothesis.

Lawrence A. Reid is affiliated with the Social Science Research Institute, University of Hawaii.
BENEDICT’S HYPOTHESIS

The New Alignment

When Benedict first proposed (1942) that the Thai languages and “Indonesian Languages” were genetically related, he did it on the basis of comparative work that he had been doing on a group of relatively unknown languages, which he concluded were genetically related to each other, and which he labelled Kadaï. These languages (Li—the speakers of which, according to Benedict, call themselves Dai—spoken on Hainan Island; and Laqua and Lati spoken in the border area of China and Vietnam) had been noted by early researchers as bearing certain similarities to the languages of Formosa, and (especially in the case of Laqua) also to Cham. It was noted that these languages had similar grammatical structures. They all had Verb-Object word order, and modifiers followed their head words, that is, Adjectives followed Nouns, Possessors followed Possessed, and so on. In addition they had noun prefixes (Lati m- and a-, Kelao hu-, Laqua ka-), which are phonetically the same as (or similar to) those commonly found in Austronesian languages.

Benedict examined the lexical material and offered about 36 proposed cognate sets, ten of which were numerals, the pronoun “I,” a demonstrative “this,” while the remainder were basic lexical items (sun, rain, water, flower, fire, man, father, head, hair, eye, ear, tooth, foot, breast, blood, fat, eat, night, weep, die, alive, black, yellow, and small). He then introduced reconstructed Proto-Tai forms into the cognate sets, adding a number of other basic lexical items such as bird, bone, sour, blind, and grandfather.

Benedict contrasted the kinds of comparisons that he was making with those that suggested a relationship between Chinese and Thai. Of the latter, there were numerals (3-10, 100) and a few body parts (such as arm, leg, and palm of the hand), but the majority were terms for animals or birds and associated cultural items (such as horse, saddle, elephant, tusk, bull, cow, hare, fowl, bee, goose, pigeon, owl) and items of trade (such as silver, indigo, ink, playing cards, salt), all of which suggest something other than a genetic relationship between Thai and Chinese. It was in this paper that Benedict first discussed the problems inherent in comparing basically monosyllabic lexical items carrying tone, which occur in the Thai and Kadaï languages, with possible cognates in Austronesian languages which are disyllabic or in some cases even trisyllabic.

It is generally recognized (Matsumoto 1973) that Thai and the other tonal languages of Southeast Asia have undergone extensive phonological attrition over the centuries and that the introduction of tone (as a part of the phonological system) was a way to maintain lexical distinctions, as certain final consonants were being eroded.

It is these phonological developments that are at the heart of the problems in Benedict’s methodology. In order to establish cognature, he must sometimes compare a Thai or Kadaï monosyllabic form with the initial syllable of an Austronesian disyllabic form (sometimes with the first syllable plus initial consonant of a second syllable in the case of CVCVC forms), and sometimes with the second syllable, for example:

<table>
<thead>
<tr>
<th>PAN</th>
<th>PAN</th>
<th>PAN</th>
<th>PAN</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>wélê</em></td>
<td><em>mcâ</em></td>
<td><em>p̥sâuq</em></td>
<td><em>bûta</em></td>
</tr>
</tbody>
</table>
| L. du | PT *ta* | PT *pot* | PT *bott*
| “eight” | “eye” | “lungs” | “blind” |

There can be no question that contrastive stress (which Benedict implies must have been present in the parent language of his proposed Austro-Tai family) has played a significant role in the phonological development of many Philippine languages. It was also
present in the parent language of this group (Zorc 1972, 1978), but the presence of contrastive stress in Proto-Austronesian is not universally accepted, despite attempts by Zorc (1983) and Dahl (1982) to prove that it was. It is perhaps significant, however, that the position of stress in the four forms cited above just happens to agree with its position in cognates in Philippine languages which still retain contrastive stress, for example, Bontok waló 'eight', matá 'eye', piáu 'heart', Binukid Manobo báu 'blind' (Reid 1971). This agreement in stress position is by no means universal, however, as one should expect given the immense period of separation between the languages. That there is any agreement is important. What needs to be done is a statistical measure to determine its significance. Nevertheless, to some linguists, Benedict's penchant for choosing either ultima or penultima in his cognate search, depending on phonetic similarity and using stress in the proto-language as an explanation, is not methodologically sound since stress in Proto-Austro-Tai cannot be independently reconstructed. Moreover, the received opinion has been that early stages of Austronesian did not have contrastive stress.

The second major problem that Benedict faced, that of reconciling forms that were often only partially phonetically similar, is dealt with only tangentially in his 1942 paper. In this he chose to rely on the presentation of his proposed cognates to establish his thesis, but did not attempt any systematic reconstruction of the parent language.

The Austro-Tai Papers

In 1966 and 1967 Benedict published a series of three articles which developed more fully the thesis introduced in 1942. These are the articles in which the term "Austro-Thai" first appeared. He stated that the stimulus for these studies came from an examination of a considerable amount of new language material that had become available to him. This material included a number of studies of the so-called Miao-Yao languages, some publications by Fang-kuei Li on the Kam-Sui languages of south-central China, and some old word lists of the Ong-Be languages spoken on Hainan. These languages are all claimed by Benedict to be part of the Austro-Tai group and to have provided him with large numbers of new cognate sets, many of which are discussed in the papers, and most of which appear in the glossary of his 1973 book.

In addition to the evidence he adduced from mainland languages, he examined material from a number of Formosan languages and claimed to have found a wealth of forms there that are directly relatable to the mainland languages. He also took note of works by Haudricourt (1951) and Goodenough (1962) which supported the reconstruction of labio-velar consonants ("kw, etc.") in Oceanic and which he felt "provided a test case . . . for evaluating the claim that Thai and Kadai are directly related to this ancestral AN language [i.e. Proto-Oceanic]."

It is in these papers that Benedict begins the systematic reconstruction of Proto-Austro-Tai. We also get a clearer picture of his methodology for dealing with the problem, mentioned at the end of the last section, of what to do with possible cognate sets that are only partially phonetically similar. One method is extension of his ultimate or penultimate syllable equation: setting up stress doublets in the proto-language, i.e., pairs of forms differing only in stress, so that a form in one language can be derived from one of the pairs, while a form in the other can be derived from the other. Another method he uses is reconstructing complex consonant sequences in the proto-language, which were reduced in different ways in the various daughter languages.

An example of both methods is illustrated by his reconstruction of the Proto-Austro-Tai word meaning 'louse' (1973:22, 333):
IN  *kutá
PT  *tthrow
Sui  tu
Mak  tau
Then  tâu
Li  sau-su
Ong-Be  kat

The first six forms Benedict derives from a PAT-reconstruction with stress on the final syllable. The Ong-Be form he derives from an alternate form which would have had stress on the penult. The presence of s in the Li form leads him to reconstruct a medial sequence, PAT *ktiu. He was able to find one other Li cognate set which supported this reconstruction, PAT *talú 'three', which, he suggests, became Li sáw-su (after loss of the initial vowel). To an Austronesianist accustomed to the extremely limited range of consonant clusters in Proto-Austronesian (even medial nasal-stop clusters in PAN are suspect, Reid 1982), Austro-Thai reconstructions look weird indeed, with syllabic initial clusters of stop/nasal + /l/ combining with pronominalization. This does not mean, of course, that such forms could not have occurred. The assumption that is problematic is that dissimilar correspondences are usually the result of different developments of proto-language clusters. Too little is known of the phonological development of most of the languages being compared to be able to make this assumption.

It is true that many of the non-Austronesian mainland languages do have very complex consonant clusters, some of which are of the type postulated by Benedict for PAT. However, it is just as likely that such clusters are the result of the reduction of disyllabic forms through the loss of unstressed vowels, or the result of prefixation, or other poorly understood processes (e.g., those which produced the initial voiceless nasals of Proto-Tai, Miao-Yao and other languages), than that these clusters were present in PAT. Benedict (1975:233) proposes, for example, a PAT reconstruction *[mə]nlək (from which the well-known PAN *manuk 'bird' would have derived) to account for the following set:

<table>
<thead>
<tr>
<th>Language</th>
<th>Reconstructed Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT</td>
<td>*nok</td>
</tr>
<tr>
<td>Diao</td>
<td>r3k</td>
</tr>
<tr>
<td>Sek</td>
<td>nok</td>
</tr>
<tr>
<td>Proto-KS</td>
<td>*{mlək</td>
</tr>
<tr>
<td>Sui Mak</td>
<td>nok</td>
</tr>
<tr>
<td>Then</td>
<td>n3k</td>
</tr>
<tr>
<td>Kam</td>
<td>nok</td>
</tr>
<tr>
<td>Lk</td>
<td>mlək</td>
</tr>
<tr>
<td>OB</td>
<td>nok</td>
</tr>
<tr>
<td>Lq</td>
<td>nok</td>
</tr>
<tr>
<td>Laha (Tu)</td>
<td>ma/nək</td>
</tr>
<tr>
<td>Laha (BB)</td>
<td>nok</td>
</tr>
<tr>
<td>Proto-MY</td>
<td>*n3?</td>
</tr>
<tr>
<td>Proto-Miao</td>
<td>*nəŋ</td>
</tr>
<tr>
<td>Proto-Yao</td>
<td>*nəŋ?</td>
</tr>
</tbody>
</table>

Benedict reconstructed the medial *ml cluster to account for two apparently aberrant forms, Kam nək and Lk mlək. But considering the fact that ml appears to be retained in the Laha form, it is just as likely that: a reconstruction such as *manuk or *mən.iək (with the medial consonant ambiguous, or indeterminate) could be made, with the Lk form maintaining the original consonant of the root.

Benedict was delighted to discover that labio-velars were present in some Oceanic languages, because he then felt justified in positing them also for PAT. Whether or not it is
possible to reconstruct them at this time depth (36000 B.C.) for PAT, no evidence has ever been produced to suggest that complex consonants of this sort were present in PAN. To the contrary, the Oceanic evidence indicates fairly clearly that such clusters were innovations in Proto-Oceanic. Furthermore, with our present subgrouping assumptions (see section on other explanations below), in which Oceanic is fairly far down on the subgrouping tree, one would need to assume that the Oceanic labio-velar consonants were independent innovations in that branch rather than retentions from Proto-Austronesian that were lost everywhere else in the family.

EVALUATING THE EVIDENCE

Recognizing that there are serious problems in the way Benedict has chosen to reconstruct PAT does not mean that the languages under discussion are not genetically related. Proving a genetic relationship is a matter of degree. Usually required are sets of sound correspondences supported by convincingly large bodies of lexical forms. The more recent the linguistic split, the easier it is to prove genetic relationship. Conversely, the more remote the split, the less likely it is that such evidence can be produced. The greater the time depth the greater the number of phonological changes that can take place obscuring forms that are cognate, and the greater the chance that cognate forms are replaced. The great number of items that Benedict reconstructs, and often with meanings of a highly specific nature, casts doubt on the validity of the work, given the great time depth that must be involved.

Having said this, I would like to take a look at some of the items that he cites, and a few others as well, and to suggest that the similarities we find are of such kinds and in such quantities that they are highly unlikely to be accidental, and probably point to a genetic relationship.

The forms I wish to discuss are those that are part of the basic lexicon (sometimes called the core vocabulary) of a language. Such forms are generally believed to be more retentive (that is, less likely to be replaced by competing forms, and less likely to be replaced by borrowing than other forms), and therefore more likely to represent true cognate sets. In addition I will discuss some morphological items, that is, forms having grammatical function or forming part of closed classes such as pronouns and demonstratives. Such items are highly retentive, lingering on in the resources of a language, sometimes with altered functions, but often relatable to similar forms in distantly related languages.

The Tai-Kadai, Miao and Yao forms are mostly taken from Benedict (1975), for which see abbreviations. The Bontok (Reid 1976), Ilokano (Carro 1957), Tagalog (Panganiban 1972) and other Philippine language forms (Isneg, Batak, and Tagbanwa) (Reid 1971) are included in order to exemplify reflexes of the Proto-Austronesian reconstructions and to demonstrate the position of stress on these forms.

Possible Cognate Sets in the Basic Lexicon

1. AN: Ben
   TK:  bulan
   bılan
   *bun
   month, moon
2. AN:  *bun
   TK:  *tun
   *tun
   mist
   rain (esp. fine rain)
<table>
<thead>
<tr>
<th>AN</th>
<th>TK</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bon</td>
<td>'kapuy</td>
<td>fire</td>
</tr>
<tr>
<td>Kam</td>
<td>'apāy</td>
<td>man, male</td>
</tr>
<tr>
<td>Sui</td>
<td>?wē</td>
<td>male</td>
</tr>
<tr>
<td>Lari</td>
<td>?wi, vui, vui</td>
<td>(gr.) child, yg. man</td>
</tr>
<tr>
<td>Bon</td>
<td>lakī</td>
<td>man (male)</td>
</tr>
<tr>
<td>Dīoī</td>
<td>lakī</td>
<td>head</td>
</tr>
<tr>
<td>Lakkia</td>
<td>laŋ</td>
<td>ru, hua</td>
</tr>
<tr>
<td>SW Thai: Ahom</td>
<td>laŋ</td>
<td>ru</td>
</tr>
<tr>
<td>C Thai: Nuag</td>
<td>?ā</td>
<td>eye</td>
</tr>
<tr>
<td>Laqua</td>
<td>?ā</td>
<td>mat</td>
</tr>
<tr>
<td>Si, Po-ai</td>
<td>ta;</td>
<td>da</td>
</tr>
<tr>
<td>Sui Mak</td>
<td>?da</td>
<td>da</td>
</tr>
<tr>
<td>Then</td>
<td>ta</td>
<td>sa, sa</td>
</tr>
<tr>
<td>Kam</td>
<td>da</td>
<td>tsa</td>
</tr>
<tr>
<td>Ong-Be</td>
<td>da</td>
<td>tsa</td>
</tr>
<tr>
<td>Li</td>
<td></td>
<td>tsa</td>
</tr>
<tr>
<td>Dukili</td>
<td></td>
<td>tsa</td>
</tr>
<tr>
<td>'maCa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si, Po-ai</td>
<td>tary</td>
<td>tai</td>
</tr>
<tr>
<td>Sui Mak</td>
<td>tai</td>
<td>dai</td>
</tr>
<tr>
<td>Then</td>
<td></td>
<td>'day</td>
</tr>
<tr>
<td>Kam</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ilk</td>
<td>'0jipan</td>
<td>tooth</td>
</tr>
<tr>
<td>TK</td>
<td>'0jipan</td>
<td>ven</td>
</tr>
<tr>
<td>Lao</td>
<td>'yan</td>
<td>phyan</td>
</tr>
<tr>
<td>Kam</td>
<td>'yan</td>
<td>phyan</td>
</tr>
<tr>
<td>Sai</td>
<td>yyan</td>
<td>phen, fen</td>
</tr>
<tr>
<td>S Li</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hova</td>
<td>'nanu</td>
<td>breast (Benedict)</td>
</tr>
<tr>
<td>SEP</td>
<td>'nanu</td>
<td>breast, milk</td>
</tr>
<tr>
<td>TK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lao</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>YHN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN</td>
<td>'kan, 'ka?an</td>
<td>eat</td>
</tr>
<tr>
<td>SW Thai</td>
<td>kin</td>
<td></td>
</tr>
<tr>
<td>N Thai: Wu-Ming</td>
<td>kin, kī</td>
<td></td>
</tr>
<tr>
<td>Dīoī</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Po-ai</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Li</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lq</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N KI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lt</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
11. AN: Bon
   TK: LK OB S Li
   *'[ID]janum
danum
  *'nam
  num
  nam, nən
  nom

12. AN: Bon
   TK: SW Thai C Thai Mak S Li N Li
   *DoDəm
dedəm
dam
dəm
  *dam
  ədəm
  dam
  ədəm
  *kəjəw
  əlgəw
  ʂədaw
  traw
  ʐaw
  *lañguy
  lañguy
  *loŋy
  liŋ
  lei
  *bəna
  low-lying
  (flooded) land
  *na:
  nən
  na

13. AN: Ilk
    TK: KS Li
    *lañguy
    lañguy
    *loŋy
    liŋ
    lei
    *bəna
    low-lying
    (flooded) land
    *na:
    nən
    na

14. AN: Ilk
    TK: OB S Li
    *manuk
    manuk
    nək
    nək
    nək
    nək
    nək
    *aLək
    ənələk
    ələk
    lik
    lik
    lak
    la:k
    *qə-łam
    ʔaɬəm
    ɬəm
    Lum
    Lum
    hum
    sum
    sem
    them
19. AN:  
   Bon  
   TK:  
   Thai  
   OB  
   WS Li  
   WS Lei  
   *bayi  
   bâ?i  
   *me:  
   mai  
   mai  
   pai  
   bai  
   *inay  
   ?inâ  

female, mother  
female  
mother  
 pref. for young girls  
or young women  
female  
woman, wife  
older sister  
mother  
mother (relationship term)  

20. AN:  
   Bon  
   TK:  
   Thai  
   Sui  
   Mak  
   Then  
   *aut  
   ?ut?it  
   *tot  
   tat  
   tut  
   *tut  
   bai  

fart  

21. AN:  
   Bon  
   TK:  
   Sui Kan  
   Mak  
   OB  
   WS Li  
   *mr?l  
   ?-um-nil  
   *hma  
   ma  
   ma  
   ma  

come  

22. AN:  
   Bon  
   TK:  
   Thai  
   Dao  
   Sek  
   *datâj  
   datâj  
   thîj  
   thîj  
   taq  
   *taq  

reach, arrive  

23. AN:  
   Bon  
   TK:  
   SW Thai  
   C Thai  
   N Thai  
   OB  
   *qabâRa?  
   ?abâga  
   *?ba  
   ba  
   va  
   ha  
   "va  

shoulder  

24. AN:  
   Isneg  
   TK:  
   Si  
   Sek  
   Mak  
   Li  

Possible Cognate Sets in the Functional Morphology  

25. AN:  
   Bon  
   TK:  
   Si  
   Lao  
   Dovi  
   Khanti  
   OB  
   S Li  
   N Li  
   Laga  
   Lati  
   MY: Miao  
   "-ku 
   -ku  
   ku  
   ku  
   ku  
   kaw  
   hao  
   hau, du  
   bo  
   khau  
   ku  
   ku  

lug pronoun  
lug pron (superior to inferior)
26. AN: White Thai
   TK: Khiamti
   Ahom
   Si
   *-Su
   su, mit
   su
   su, t
   su:
   2sg pron
   2sg pron (superior to inferior)

27. AN:
   TK: W Li
   N Khoao
   *-mu
   mo
   mu
   2pl pron

28. AN:
   TK: Tag
   Si, Lao, BT, WT, Lk
   *ni
   ?ini
   ni
   thou, you
   this

29. AN: Bon, Ilk
   TK: Nung, Lk, KS
   Si, Li
   Li
   MY: Yao
   nay
   nai
   nei
   nai
   here, this
   this

30. AN:
   Bon
   TK: Li
   Loi, WS Li
   Si, Lao
   Lk
   MY: Yao
   Miao
   *na
   na
   nan
   nan
   nan
   nan
   na
   that; ligature; determiner
   this; -ns 3s pr; determiner
   3s
   that, there
   that (near addressee)
   that
   3s pron

31. AN:
   Tag
   TK: Lk
   Lq
   Li
   MY: Miao
   *tu
   ?itu
   tu
   to
   to
   that
   they
   that
   that far (genitive)
   that far

32. AN:
   Tag
   TK: Si
   MY: Miao
   *di
   di
   *di:
   thi:
   ti:
   that
   that (far); locative
   that (far)
   place; locative

33. AN:
   Bon
   TK: SW Thai
   C, N Thai
   *ti
   ti
   that
   determiner
   that (far)

34. AN:
   Ilk
   Batak
   Tagbanwa
   MY: Miao
   ti
   ti
   ti
   ti
   negative

35. AN:
   MY: Miao
   "Di
   ti
   negative

A third set of possible cognates is found in the numerals, a number of which are very similar in form to those found in Austronesian languages. Benedict (1975:444) provides a chart of numerals from a number of languages, and discusses those that he believes are cognate.
OTHER EXPLANATIONS

There are a variety of explanations, other than postulating a genetic relationship, for the similarities that are discussed above and that have been noted by so many other linguists. One class of explanations may be categorized as language contact explanations. These include borrowing, substrata (and other kinds of stratal influence), and areal influence or diffusion. These are the kinds of explanations that are paramount in accounting for many of the similarities found between Thai, Vietnamese, and Chinese, including their monosyllabic syllable structure and tone. They all assume a period of linguistic contact, varying from intermittent and casual trading relationships to extended periods of geographical contiguity resulting in “diffusion” of features between languages, and periods of contemporary occupation of the same geographical territory resulting in “stratal” influence. Both of the latter types of contact imply extensive bi- or multilingualism.

In the context of the present discussion we need to ascertain the degree of likelihood that one or more of the above relationships existed between the ancestors of the Austroasiatic languages and the ancestors of the Thai languages. At the present time there are only two Austroasiatic groups on the mainland: the Chamic languages, and Malay and its related languages (such as Urak Lawoi’ which is spoken on Phuket in Thailand). Neither of these groups gives any evidence that they are residual enclaves of some Austroasiatic homeland. Both Malay and the Chamic group can be unambiguously assigned to a Western Malayo-Polynesian subgroup. Virtually all Austroasiaticists believe that the ancestors of the Chamic group moved back to the mainland from some area within the Malay-Indonesian language area. We have no evidence moreover that the Chamic group on the mainland has ever been in contact with other than the Austroasiatic languages (Khmer, Bahnaric and Viet-Muong) that presently surround it. And although the Malay group presently is contiguous to the languages of southern Thailand, we know that this contact is relatively recent—within the last 800 years or so—as Thai speakers moved from the South China area and split the Mons on the Burmese side from their Austroasiatic Khmer cousins on the Kampuchean side.

To establish areal diffusion (or stratal influence) as a likely explanation, it would be desirable to have evidence that there was an Austroasiatic homeland on the mainland, and preferably in the South China area. But we know of no languages spoken in this area that are unambiguously Austroasiatic, let alone one which would show evidence of representing an Austroasiatic homeland. If there had been one in the past which has now been sinicized or for some other reason has disappeared, or if the Pre-Austroasiatic ancestors of the Proto-Austroasiatics lived in this area on the mainland (as indeed is probable), the possibility of language contact as an explanation for the similarities we are discussing would exist.

Borrowing as a result of some kind of trading relationship is unlikely, for at least two reasons. First, the kinds of terms we discussed above are not the kind that are likely to be borrowed in such a contact situation. Second, we have no other evidence that early Austroasiatics and early Thais carried on such trading, although they may have. If Pre-Austroasiatics moved from Mainland Southeast Asia to what is now Taiwan they could also have returned. But at the time depth we are talking about, which must have been prior to the dissolution of the Proto-Tai community, there is no evidence that such trade occurred.

The other kinds of possible explanations are of two types. The first accounts for similarities which are the result of the inherent character of language. Such an explanation is
often given for agreement in syntactic patterning which is not infrequently found between genetically unrelated languages. Such agreement could be considered supportive of a genetic relationship based on other criteria but may not of itself establish the relationship. The fact that Thai and Indonesian are both SVO tells us nothing about their genetic relationship. Chinese is also SVO. Thai and its related languages do have a head-modifier word order, like Proto-Austronesian and most Austronesian languages, with adjectives following their head nouns, and possessor nouns following possessed nouns, unlike Chinese. But explanations of this sort do not account for the similarity in lexical forms that has been discussed above.

The other explanation that has been used to account for these similarities is coincidence. It is proposed that because of the limited number of phonemes in language and their limited combinatorial possibilities, accidental similarities are bound to occur, and therefore coincidences of the type Thai fai—English fire, Thai tai—English die, Thai rim—English rim, must be expected and do not suggest genetic relationship. This is true, but it would be extremely improbable for coincidence to bring about such striking similarities, not only in the core vocabulary but in the morphology as well.10

AUSTRONESIAN SUBGROUPING AND THE HOMELAND HYPOTHESES

Postulating an Austro-Thai genetic relationship, or even a close contact relationship, has implications for an Austronesian homeland. It can probably be assumed that homelands were located in the geographical vicinity of the “seams” between first order subgroups. If Austronesian and Thai are genetically related, then postulating Formosa as the Austronesian homeland as Benedict and others have done is reasonable, because of its geographical contiguity to Southeast China, the presumed homeland of Proto-Tai. Even if they are not genetically related, and if the similarities between the groups are the result of extensive linguistic contact, we would still need to place pre-Austronesians in Southeast China.

It may be useful at this point to review the various hypotheses regarding the homeland of the Proto-Austronesians, considering the degree to which they are supportive of a Formosan homeland, or whether other possibilities exist. Crucial to this discussion is the position of Oceanic languages within Austronesian. Early subgrouping hypotheses divided Austronesian into two major families, a Western group—Hesperonesian, and an Eastern group—Oceanic. Such a subgrouping is explicit in the work of Tsuchida (1976), who makes Formosan languages a branch of the Hesperonesian group. Haudricourt’s (1965) subgrouping is similar to Tsuchida’s, but with Formosan made a third primary branch. Dyen’s (1965) lexicostatistical subgrouping had a large number of “highest order subgroupings” (24 of them simple languages, 12 of them small groups), the majority of which were clustered in three areas; Melanesian–East New Guinea, West New Guinea, and Formosa. Dyen suggested each of these areas as a possible homeland, based on the assumption that areas of greatest linguistic diversity implied longest periods of settlement. Dyen (1964) subsequently removed Formosan from the list of possible homelands because he claimed that the lexicostatistical percentages in this area were deflated. His most recent statements (Dyen 1982) continue to keep Oceanic as a primary subgroup, and he still believes (Dyen, personal communication) that the homeland was probably in one of the two areas he formerly cited.

Other linguists do not consider Oceanic languages to be a primary subgroup, thus in effect removing the Oceanic-Hesperonesian seam (that is, the area of Eastern Indonesia—
Western New Guinea–Melanesia is a possible homeland. Dahl’s (1973) subgrouping makes all non-Formosan languages a single subgroup, a position also taken by Blust (1977), who labels this subgroup Malayo-Polynesian. In Blust’s version, Oceanic languages are at the end of the line, a fifth order subgroup (see Fig. 1).

The evidence that has been produced for considering all non-Formosan languages a single subgroup is not extensive, but it is persuasive. It includes several phonological innovations found in all non-Formosan languages, plus a number of morphological innovations. The phonological innovations are as follows:

1. *t and *t’ → *t
2. *l and *L→ *l
3. *n and *–L→, –L → *n

Of these, numbers (2) and (3) represent a “split merger” that would not likely have developed independently in more than one group.

The morphological innovations in the non-Formosan languages which can be reconstructed for their parent (but not for Proto-Austronesian) include the following:
1. "paN-" (and associated "maN-" and "minaN-") transitive prefix
2. "siDa" 3 pl. Nominative pronoun
3. "-mu" 2 sg. Genitive pronoun (PAN 2 pl.)
4. "-atəm" 1 incl. pl. Predicative pronoun (PAN "-itəm")

In summary, an eastern Indonesian–western Oceanic homeland for Proto-Austronesian, which would contraindicate an Austro-Thai genetic hypothesis, is supportable only on the basis of lexicostatistics. Subgrouping hypotheses that imply a Formosan Austronesian homeland are more generally accepted and are supported by fairly persuasive qualitative evidence.

THE AUSTRIC HYPOTHESIS

First proposed by Schmidt (1906), and both supported and refuted by many linguists since, the Austric hypothesis would link Austronesian and Austroasiatic languages into one superstock. This hypothesis is relevant to the present discussion, because if it is possible to show that Austronesian and Austroasiatic languages are probably genetically related, it would enhance the possibility of an Austronesian-Thai genetic relationship because of the geographic distribution of the languages.

The similarity of some of the Austronesian morphology to Austroasiatic, especially the "pa- 'causative', has been noted in the literature. It appears, however, that the full extent of the similarities has not been recognized. They are in fact so distinctive that only a genetic explanation can adequately account for them.

Nancowry, a language classified by all who have examined it as Austroasiatic, has a morphological apparatus so similar in form and function to what has been reconstructed for PAN that an Austronesianist looking only at the morphology would immediately consider it to be Austronesian. Nancowry, spoken in the Nicobar Islands, is apparently a relic area, sufficiently isolated from the Mainland Southeast Asian linguistic area to have escaped the innovative pressures that resulted in the loss or modification of the original morphological features from many of its sister languages.

The following comparisons, based on a brief article by Radhakrishnan (1976), should be sufficient to illustrate the point.

<table>
<thead>
<tr>
<th>Nancowry</th>
<th>Austronesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>ha-</td>
<td>&quot;pa-&quot; causative</td>
</tr>
<tr>
<td>h-an-</td>
<td>&quot;paN-&quot; causative instrumental</td>
</tr>
<tr>
<td>ma-</td>
<td>&quot;maR-&quot; agentive</td>
</tr>
<tr>
<td>-in-</td>
<td>&quot;-in-&quot; completive</td>
</tr>
<tr>
<td>-a</td>
<td>&quot;-a&quot; objective</td>
</tr>
<tr>
<td>-um-</td>
<td>&quot;-um-&quot; agentive (also causative in some forms)</td>
</tr>
</tbody>
</table>

The similarities both in form and function are striking. But perhaps what is even more striking is that the process that produces infixation in the Austronesian languages is apparently still operating in Nancowry. Infixed in all of these languages appear after the
Austro-Tai

Miao-Yao

Kadai

Austro-Kadai

Austro-Japanese

Austronesian

Japanese-Ryukyuan

NOTES

1. More recently, Benedict has included the Thai languages within the Kadai family.
2. The Thai writing system, less than 700 years old, makes a number of phonological distinctions, and represents various consonant clusters no longer used in modern Thai.
3. The asterisks before the forms in this list, and elsewhere in the paper, mark lexical items which are supposed daughter languages. The equals sign (=) denotes a proposed genetic relationship between the protoform on the left (in this case Proto-Austro-Asiatic) and the form on the right of the arrow. This language may itself be a proto-language, as in the case of the Proto-Tai forms for "eye," "heaps," and "blind."

4. Although Benedict used the term "Austro-Tai," the term more commonly used today is "Austro-Asiatic" language of Bangkok, also known as Siamese, belongs.

5. Miao is perhaps better recognized today as Hmong, or Mong, the language of large numbers of refugees in the United States and elsewhere.

6. Philippine languages show penultimate stress on this root.

7. It should be noted that Benedict himself calls his cognate sets LCG's, i.e., Likely Cognate Groups having a better than 50 percent chance of being cognate (1975:139). He furthermore labels his reconstructions as "provisional... of a kind that might be labeled simply 'work in progress'" (1975:146).

8. Li (1977:262) suggests the vocalic alternation in this set "may be due to an original diphthong 'ï', but the reconstruction is doubtful."

9. Proto-Austro-Asiatic was probably VOS, or VSO, and Indonesian has only relatively recently undergone a systemic innovation which reordered the basic sentence constituents.

10. For an excellent discussion of the various possibilities discussed in this section, but with reference to the languages of Thailand, see Marisoff (1973).

**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>AT</th>
<th>Proto-Austro-Asiatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>BON</td>
<td>Bonok</td>
</tr>
<tr>
<td>BT</td>
<td>Black Tai</td>
</tr>
<tr>
<td>C</td>
<td>Central</td>
</tr>
<tr>
<td>ILK</td>
<td>Ilokano</td>
</tr>
<tr>
<td>IN</td>
<td>Indonesian</td>
</tr>
<tr>
<td>KS</td>
<td>Kau-Sui</td>
</tr>
<tr>
<td>LK</td>
<td>Lakkia</td>
</tr>
<tr>
<td>Lg</td>
<td>Laos</td>
</tr>
<tr>
<td>Lz</td>
<td>Lazi</td>
</tr>
<tr>
<td>My</td>
<td>Microwave-Asiatic</td>
</tr>
<tr>
<td>N</td>
<td>North</td>
</tr>
<tr>
<td>OB</td>
<td>Ong-be</td>
</tr>
</tbody>
</table>

**REFERENCES**

Benedict, Paul K.
N.D. Japanese/Austro-Asiatic.

Blust, Robert

Cardo, Andres

Dahl, Otto C.
DEMPOULOF, OTTO

DYN, ISIDORE

GOODENOUGH, WARD

HAUDRICOURT, ANDRÉ

LI, FANG KUEI

MITSOFF, JAMES A.

PANGANIBAN, JOSE VILLA

RADHAKRISHNAN, R.

REID, LAWRENCE A.

SCHMIDT, WILHELM

TSUCHIDA, SHIBERU

ZONG, R. DAVID