

# The Current State of Linguistic Research on the Relatedness of the Language Families of East and Southeast Asia<sup>\*</sup>

The current state of knowledge on the relatedness of the major language families of East and Southeast Asia is summarized following a major conference on this topic in Hawai'i in 1993. It is concluded that the Austric hypothesis linking the Austronesian and Austroasiatic language families is based on good evidence and the "homeland" of Austric was perhaps in the middle Yangzi Valley. On the other hand, the hypotheses suggesting common roots between Austro–Thai and Austronesian, and between Japanese and Austronesian, did not find much support and the numerous similarities can best be explained as the result of language contact. The relationship between Chinese and Austronesian was not resolved and merits more study.

## 1. Introduction

Throughout this century, there has been an ongoing examination of the possible relationships that exist between the various language families that are represented in East and Southeast Asia. Within recent years, however, there has been a burgeoning interest in this topic as more, and in many respects better, descriptions become available of the languages spoken by the enormous number of ethnic groups in the geographic areas under discussion.

In an attempt to capitalize on this interest, and the fortuitous presence in Hawai'i in 1993 of a number of the linguists involved in these endeavors, the National Science Foundation was approached with a request to fund a conference which would enable us to bring not only the major proponents of each of the main hypotheses dealing with the relatedness of these language families, but also their major critics. To complement the linguistic discussions we added in a physical anthropologist (Pietrusewsky 1994) and an archaeologist (Bellwood 1994). This Conference on Austronesian/Mainland–Asia Connections (CAMAC) was funded and held at the University of Hawai'i over a four day period in May 1993.

---

<sup>\*</sup> Originally published as: The current state of linguistic research on the relatedness of the language families of East and Southeast Asia. In *Indo-Pacific prehistory: The Chiang Mai papers*, Vol. 2, ed. by Ian C. Glover and Peter Bellwood, 87-91. Bulletin of the Indo-Pacific Prehistory Association 15. Canberra: Australian National University. (1996)

This paper will provide a brief overview of the presentations given during that conference and will give an evaluation of the present state of what can be reasonably asserted about the genetic relationships of the language families of the Indo-Pacific and adjacent areas.

We begin with the assertion that although their internal relationships still engender considerable controversy, the major language families of these areas are themselves pretty much beyond dispute. They are: Austronesian, Tai-Kadai, Austroasiatic, and Sino-Tibetan. The major purpose of the CAMAC conference was to examine the possible connections between the first of these groups, Austronesian, the great majority of whose speakers constitute the main populations of island Southeast Asia and the Pacific, and the other language families, whose speakers constitute the major populations of mainland East and Southeast Asia. In addition, we decided to look also at the possible relationship of the Austronesian family with Japanese.

## **2. Austronesian and Austroasiatic**

The Austroasiatic family as described by Parkin (1991) consists of over 150 languages,

...scattered in a considerable sweep of tropical and subtropical Asia, from Nimar (southwestern Madhya Pradesh) in the west to the Vietnamese shore of the South China Sea in the east and from Yunnan in the north to the Endau river in the south.... Austroasiatic speakers are represented in every nation state of continental Southeast Asia (Burma, Thailand, West Malaysia, Vietnam, Laos, Cambodia) as well as in southern China, India, Nepal, Bhutan, and Bangladesh.

The proposal that this family is genetically related to Austronesian, in a superfamily called Austric, has had a long and chequered career, with papers, both pro and con, having been published from late last century until the present day.

Present at the CAMAC conference were two of the current promoters, Gérard Diffloth and myself, with Paul Benedict being there as the main detractor. His position that the two families are not genetically related had appeared several times in print (Benedict 1976a, b). There were in addition several self-proclaimed skeptics, probably the most prominent of whom was Robert Blust. Diffloth and I had agreed to present complementary papers, with him presenting lexical evidence for Austric, and me focusing on morphological and

syntactic evidence. There has never been any shortage of proposed cognate sets for Austric. Schmidt (1906) had proposed 215 sets. Others have been proposed by Shorto (1976), and by Pou and Jenner (1974). The most recent contributor (Hayes 1992) claims to have identified over 750 potential cognate sets. Diffloth's paper was the beginning of an attempt to review the proposed cognates (excluding those by Hayes, whose article had not been published and was not referred to), from the point of view of recent advances in Mon-Khmer comparative reconstruction. In the process he succeeded in rejecting a considerable number of unsubstantiated comparisons, and explaining away many more. The result, Diffloth claimed, was that, at the present time,

...the lexical agreement between AA and AN... is not impressive, whichever lexical class one chooses to look at, but it is undoubtedly there. Verbs and statives constitute the majority.... For nominals, the record is weaker: we find some natural objects, a few body parts and kinship terms, even fewer artifacts and animals, practically no plants, hardly any pronouns and no numerals.

The impression, he states, is negative to those concerned with quantity [of verifiable cognates] (Diffloth 1994).

My paper (Reid 1994) was concerned with evaluating claims made about apparent similarity between the morphology of the languages of the two families. I was able to show that there are substantial areas of agreement both in the forms as well as the functions of prefixes, infixes, and at least one suffix, as well as in general structure of what must be reconstructed for the form of the noun phrase for the parent languages of both families, with agreement also in several of the case marking prepositions. I was also concerned with attempting to show that these agreements in form and function of verbal affixation could not be the result of borrowing. The result was that the Austric hypothesis now seems to be fairly firmly established.

### **3. Austro-Tai**

This hypothesis, probably better known and more widely discussed than the Austric hypothesis, was first proposed by Benedict (1942), although as early as 1901, Schlegel had proposed that Malay and Siamese were related. In its simplest form it proposed that the family of languages known as Tai-Kadai is genetically related to Austronesian. The body of evidence that has been pre-

sented to establish the Austro-Tai hypothesis has primarily been gathered and presented by Benedict. It ranges from what are really striking similarities in form and meaning between a considerable number of terms, many from the area of basic vocabulary (the forms first presented in Benedict 1942) to a large number of probably spurious comparisons found in Benedict (1975). The problem has always been that there seems to be too many good comparisons to be accounted for by chance, but not enough systematicity in the sound correspondences to unequivocally accept a genetic hypothesis as the explanation for the similarities. Under such circumstances a contact relationship is generally understood to be the explanation.

At the CAMAC conference it was just such a conclusion that was drawn by Thurgood in a carefully-argued paper, “Tai-Kadai and Austronesian: the nature of the relationship”. It is instructive that Thurgood began his preparation for the conference paper as one who was already predisposed to the Austro-Tai hypothesis. In fact, he had been invited to present a paper which would outline the best evidence for Austro-Tai. However, after looking at the evidence, he stated, “The results of that careful examination are surprisingly unambiguous: a careful examination of the currently available reconstructions in various subgroups of Tai-Kadai strongly suggests that many of the Austronesian forms usually cited as proof of a genetic relationship between Austronesian and Tai-Kadai are the result, not of a genetic relationship, but instead are the product of language contact. In short these words behave precisely as one would expect loan words to behave. In short, they are loans” (Thurgood 1994).

#### **4. Sino-Austronesian**

The theory that Chinese and Austronesian languages might be genetically related seems to be difficult for some linguists to even consider. On the surface, the complex, tonal phonology, and isolating morphology of Chinese languages seem to be vastly dissimilar from the phonologically relatively simple, inflectional morphology of many of the better known Austronesian languages. The theory that Chinese and Austronesian languages are related goes back to a proposal made by Wulff (1942). But until recent work by the French scholar, Laurent Sagart, the proposal had pretty much lain dormant. Sagart’s paper for CAMAC (Sagart 1994) was one of the most carefully researched and methodo-

logically sound presentations of the conference. Laurent was able to demonstrate several clear areas of agreement between these two proto-languages. One of these areas is in the area of word morphology. As Sagart notes, “Because [morphology] is internally strongly structured and resistant to change, both internal and external... the diagnostic value of morphological congruence in determining genetic relationships is particularly high.”

Sagart proceeded to provide a considerable number of data sets from Austronesian languages and Old and Modern Chinese languages to support his claim that at least two infixes (Proto-Austronesian *-ar-* and Old Chinese *-r-* ‘distributed object’, and the ubiquitous Proto Austronesian *-in-* ‘patient nominalizing infix’, Old Chinese *-j-* ‘non-agent nominalizing infix’), and one prefix (Proto-Austronesian *ma-*, Old Chinese *N-* ‘stative/intransitive verb prefix’) are shared between the two families. He then presented more than fifty cognate sets, carefully chosen to demonstrate full agreement with the sets of sound correspondences that he proposes between Proto-Austronesian and Old Chinese. He also required for this set clear semantic equivalence. The scholarship of the paper is impressive and the results (to me at least) persuasive, although some argued that there were too few cognate sets from the area of basic vocabulary for the general thesis of the paper to be convincing.

If Chinese and Austronesian are genetically related, then what does this say about the relationship between Austronesian and Tibeto-Burman, since most scholars would probably agree that the Sino-Tibetan hypothesis is reasonably well established (Benedict 1972). However, van Driem ([n.d.](#)) claims a rather different view of the position of Chinese in relation to the rest of the family than has usually been claimed, and Sagart addresses this point also. He says that the claim he makes of a genetic unity between Old Chinese and Proto-Austronesian is not exclusive of other links between either language and other East Asian language groups, further stating that *prima facie* evidence does exist for Tibeto-Burman. He cites apparent cognates of the *pa-* ‘causative’ prefix, and claims there is also evidence that the TB medials *-r-* and *-y-* were infixes and that some of their functions were shared with the corresponding Chinese and Austronesian infixes. He also provides a small set of interesting looking comparisons from the area of basic vocabulary (“potential cognates”) between

Proto-Tibeto-Burman and Proto-Austronesian, sufficient to suggest that a systematic comparison of the two languages might not be wasted time.

### **5. Japanese-Austronesian**

The stepchild of all these hypotheses, the one that is viewed with most suspicion by the majority of linguists, is that which claims a genetic relationship between Japanese, and either Austronesian as a whole or some subgroup of Austronesian, such as Oceanic or Polynesian languages. It is not that Proto-Austronesian was an ergative language, had verb-initial syntax, head-dependent word order, and other concomitant syntactic features which are polar opposites to Japanese accusative, verb final, dependent-head word order. That such a radical shift can take place in the history of a language is clearly indicated by the Munda languages of India, which have shifted from Proto-Austroasiatic verb-initial syntax to verb-final syntax, probably (but not entirely) due to the influence of surrounding Indic languages.

The major problem is that the numerous cognate sets that have been proposed suffer from insufficient regularity in their correspondences to be able to assert that their similarities are due to genetic inheritance. At the CAMAC conference, there were two presentations in support of the hypothesis, one by Paul Benedict whose published work (Benedict 1990) gives a full presentation of the evidence he uses to support his views. The other was by Takao Kawamoto, an elderly Japanese scholar who has worked for close to 30 years trying to establish a relationship between the two (Kawamoto 1977, 1978, 1981, 1984). Benedict, in his inimitable way used his whole panoply of neatly-labeled explanations for otherwise quite incomparable forms, and built a castle that looks impressive until the foundations upon which it is built are closely examined. Kawamoto's paper consisted primarily of sets of words that bore some semantic and phonetic similarity, between Japanese and one or another language or subgroup within Austronesian. He admitted the problem of inadequate sound correspondences, proposing three 'waves' of Austronesianization of the Japan Islands, the terminology implying that he really is not talking about a genetic relationship at all, but a contact one, which is at least to me, highly plausible. Is it conceivable that the great voyagers only headed their canoes south and east from Formosa, and never ventured north to the islands which were only a few

days sailing away from their own shores? The details of Kawamoto's proposals are somewhat implausible, so I shall not reiterate them here, but contact as an explanation for at least some of the similar forms that he cites is quite plausible.

The contrary position was presented by Alexander Vovin, a specialist in the reconstruction of the Altaic language family, which includes Korean, Mongolian, Manchu-Tungus and Turkic, and to which he (and probably most other linguists) believes Japanese also belongs. He presented convincing arguments that many of the apparent cognates proposed by Benedict between Old Japanese and Proto-Austronesian no longer appear to be cognate when the earlier, Proto-Japanese, forms are taken into consideration, and that for many of these forms their Altaic etyma are much better supported (Vovin 1994). The overall impression, then, is that Japanese is probably an Altaic language, as has been long claimed, but that it shows the effect of probable Austronesian influence at some early stage in its history.

## **6. Conclusion**

At this point then the Austric hypothesis, linking the Austronesian and Austroasiatic language families is probably the best supported of the various hypotheses. Robert Blust, a self-proclaimed skeptic about all of the proposals of long range genetic relationships with Austronesian, states that after seeing the present evidence for Austric, his skepticism, "crumbled rapidly. Here at last was a claim about the external relationships of the AN languages that appeared to meet the same demands made of established language families: that chance and borrowing be convincingly eliminated as alternative explanations of observed similarity before advancing a hypothesis of genetic relationship" (Blust 1996).

Finding a homeland for this superfamily is no simple task, in that the homeland of Austroasiatic was probably in the Burma–Yunnan border area, and the homeland of the Austronesian family was most probably in Taiwan. The middle Yangzi region, with its ancient record of rice cultivation could have been the homeland. Blust, however, looks further west, to the Upper Burma–Yunnan region itself, and suggests that archaeological research in this area might show it to be a center for agricultural innovation. He goes further and suggests that down-river migration of pre-Austronesian speakers would have brought them

past what was probably the ancestral home of the Tai-Kadai languages, the valley of the Wu tributary of the Yangzi River in Guizhou Province.

Since Blust has not accepted the Sino-Austronesian hypothesis, he finds no need to bring Chinese, or more broadly the Sino-Tibetan languages into his calculations. At least one other proposal exists, that considers the possibility that all the families (excepting Japanese) are genetically related, and makes a tentative claim for a proto-language, Proto-East Asian, with homeland between the Yangzi and Yellow Rivers, and from which each of the families of the area developed (Starosta pers. comm.; see also Bellwood 1994).

### References

- Bellwood, Peter. 1994. An archaeologist's view of language macro-family relationships. *Oceanic Linguistics* 33 (2):391-406.
- Benedict, Paul K. 1942. Thai, Kadai and Indonesian: A new alignment in Southeastern Asia. *American Anthropologist* 44:576-601.
- Benedict, Paul K. 1972. *Sino-Tibetan: A conspectus*. London: Cambridge University Press.
- Benedict, Paul K. 1975. *Austro-Thai: Language and culture*. New Haven: HRAF Program.
- Benedict, Paul K. 1976a. *Austro-Thai and Austroasiatic*. In *Austroasiatic studies*, Part I, ed. by Philip Jenner, Laurence C. Thompson, and Stanley Starosta, 1-36. Oceanic Linguistics Special Publication No. 13. Honolulu: University Press of Hawai'i.
- Benedict, Paul K. 1976b. Shorto: In defense of Austric—Comment. *Computational Analyses of Asian and African languages* 6:105-8.
- Benedict, Paul K. 1990. *Japanese/Austro-Tai*. *Linguistica Extranea, Studia* 20. Ann Arbor: Karoma Publishers, Inc.
- Blust, Robert. 1996. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. In *Prehistoric settlement of the Pacific*, ed. by Ward H. Goodenough. Transactions of the American Philosophical Society 86.5:117-140, Philadelphia, American Philosophical Society.
- Diffloth, Gérard. 1994. The lexical evidence for Austric, so far. *Oceanic Linguistics* 33 (2):309-22.

- Hayes, La Vaughn H. 1992. On the track of Austric: Part I. *Mon-Khmer Studies* 21: 143-78.
- Kawamoto, Takao. 1977. Towards a comparative Japanese-Austronesian I. *Bulletin of the Nara University of Education* 26 (1):23-49.
- Kawamoto, Takao. 1978. Towards a comparative Japanese-Austronesian II. *Bulletin of the Nara University of Education* 27 (1):1-24.
- Kawamoto, Takao. 1981. Proto-Oceanic paradigms and Japanese. *Bulletin of the International Institute for Linguistics Sciences, Kyoto Sangyo University* Vol. II (4): 139-55.
- Kawamoto, Takao. 1984. Two sets of sound laws between Japanese and Austronesian. *Bulletin of the Joetsu University of Education* 3:31-50.
- Parkin, Robert L. 1991. *A guide to Austroasiatic languages and their speakers*. Oceanic Linguistics Special Publication No. 23. Honolulu: University of Hawai'i Press.
- Pietruszewsky, Michael. 1994. Pacific-Asian relationships: A physical anthropological perspective. *Oceanic Linguistics* 33 (2):407-29.
- Pou, Saveros, and Philip Jenner. 1974. Proto-Indonesian and Mon-Khmer. *Asian Perspectives* 17:112-24.
- Reid, Lawrence A. 1994. Morphological evidence for Austric. *Oceanic Linguistics* 33 (2):323-44.
- Sagart, Laurent. 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic Linguistics* 33 (2):271-308.
- Schlegel, G. 1901. Review of Frankfurter's Siamese grammar, *T'oung Pao* 2:76-87.
- Schmidt, Wilhelm. 1906. *Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens and Austronesiens*. Archiv der Anthropologie, Braunschweig, n.s. 5:59-109.
- Shorto, Harry L. 1976. In defense of Austric. *Computational Analyses of Asian and African Languages* 6:95-104.
- Thurgood, Graham. 1994. Tai-Kadai and Austronesian: The nature of the historical relationship. *Oceanic Linguistics* 33 (2):345-368.
- van Driem, George. 1995. Black Mountain conjugational morphology, Proto-Tibeto-Burman morphosyntax, and the linguistic position of Chinese. In *New Horizons in Tibeto-Burman Morphosyntax*, ed. by Yoshio Nishi, James

- Alan Matisoff and Yasuhiko Nagano, 229-259. *Senri Ethnological Studies* 41. Osaka: National Museum of Ethnology.
- Vovin, Alexander. 1994. Is Japanese related to Austronesian? *Oceanic Linguistics* 33 (2): 369-90.
- Wulff, K. 1942. *Ueber das Verhaeltnis des Malayo-Polynesischen zum Indo-chinesischen*. Det Kgl. Danske Videnskabernes Selskab. Historisk-filologiske Meddlsr XXVII, 2. København: Ejnar Munksgaard.