

On the Track of Austric: Part I. Introduction

La Vaughn H. HAYES

Abstract

Austric, the superstock which would aggroup the Austroasiatic and Austro-Tai languages of South East Asia, has been controversial since Wilhelm Schmidt first proposed it in 1906. In one of its more influential critiques, Paul K. Benedict advised in 1973 that the hypothesized grouping be abandoned due to absence of significant agreement in the core vocabulary. Since 1983, however, the writer has determined that such agreement does exist and in quantity commensurate to that asserted by Benedict to occur amongst the primary subgroups of his Austro-Tai stock.

In this paper, the first in a short series purporting to present the writer's new evidence for Austric, the crucially probatory problems for the proposed genetic relationship are discussed, some new aspects of Austroasiatic linguistic history are briefly introduced, and some of the lexical, phonological, and morphological evidence supporting the Austric hypothesis is presented.

On the basis of this discussion and presentation, it is found that Austroasiatic and Austro-Tai do share partial correspondence indicative of common origin, including agreement in the core vocabulary. Consequently, it is also concluded that these language stocks are genetically related and Schmidt's Austric hypothesis can at last be acknowledged as a valid concept for areal language classification.

1. Introduction.¹

1.1 Purpose and Objectives

This paper serves as the introduction to a serial study, "On the Track of Austric," the general purpose of which is presentation of new lexical evidence supporting the verification of the Austric hypothesis. The primary objective of this paper is to

¹ The bulk of the AA lexical data used in this study was collected in the field by members of the former Vietnam Branch of the Summer Institute of Linguistics, often under arduous and hazardous conditions. I would like to express my thanks and appreciation to them, as well as the other linguists and staff of this organization who have made that data available to armchair researchers such as I, for without their efforts, none of this would be possible.

introduce the reader to the hypothesis, briefly review past efforts to verify it, and discuss the probatory problems which must be overcome in any attempt to verify it.

The secondary objective of this paper is to make a contribution to setting the stage for reconstruction of the Austroasiatic (AA) and Austric proto-languages. This purpose will be served by introducing some important new aspects of AA linguistic history and presenting the proposed phonological and morphological systems of those proto-languages.

1.2 *The Austric Hypothesis*

In 1906, Wilhelm Schmidt established AA as a superfamily composed of the Munda, Khasi, Nicobar, and Mon–Khmer families.² The some 140 languages of this stock are located in eastern India, the Nicobar Islands, and throughout the Indochina peninsula.³ At the same time, he proposed the existence of an Austric superstock comprising AA and Austronesian (AN).⁴ The some 700 languages of the latter superfamily, known as Malayo–Polynesian until Schmidt renamed it in 1899, are found in a vast geographic area in insular South East Asia and most of Austronesia, with outliers on Madagascar, Malaysia, Vietnam, Hai Nan, and Taiwan. In support of his Austric grouping, Schmidt cited 215 lexical comparisons between AA and AN and highlighted certain congruences in their sound systems, methods of word formation, and grammatical structure.

Over the ensuing 86 years, Schmidt's far-reaching proposal has been periodically argued back and forth with largely inconclusive results. No one has been able thus far to adduce really convincing evidence that the Austric superstock does or does not exist. Only five of the more recent discussions of the hypothesis will be mentioned here.

In 1975, Saveros Pou and Philip N. Jenner sought to reopen the debate by focussing attention on a possible cognate relationship between MK and AN. They presented a variety of correspondences and listed 65 possible cognates, but felt it premature to voice any conclusions about the indications of their findings, the primary reason being an admitted lack of sufficient data permitting establishment of reliable sound correspondences.

Also in 1975, the results of Paul K. Benedict's examination of the Austric question were published as Appendix II, "Austro–Thai and Austroasiatic", to his book, *Austro–Thai: Language and Culture, with a Glossary of Roots*.⁵ Benedict demonstrated that AN and two language families of mainland South East Asia,

² Schmidt initially included Chamic, an AN subgroup in Vietnam.

³ Cf. Parkin 1991 for a recent listing of the AA languages.

⁴ Other abbreviations used in this paper are AJ (Austro–Japanese), AK (Austro–Kadai), D (Dempwolff), EAA (Eastern AA), FO (Formosan), L (Lopez), MK (Mon–Khmer), MUK (Muong Khen), PAA (Proto–AA), PAT (Proto–AT), PM (Proto–Mon), PMK (Proto–MK), PMN (Proto–Mnong), PMY (Proto–Miao–Yao), POC (Proto–Oceanic), PS (Proto–Semai), PW (Proto–Waic), VN (Vietnamese).

⁵ The appendix is a revised version of a paper which was presented at the First International Conference on AA Linguistics held at Honolulu, Hawaii, in January 1973 and published with other communications to the conference in *Austroasiatic Studies* in 1976.

Kadai and Miao-Yao, are genetically related and constitute the AT stock (he has since added Japanese-Ryukyuan, cf. Sec. 2.2.1). As a consequence, Austric—if it exists—would have to comprise AA and AT.

In his approach to the Austric problem, Benedict (1975:464f.) attempted to answer this basic question:

In addition to the congruency in over-all configuration, which might be ascribed to areal factors, do Austro-Thai and Austroasiatic share a common corpus of roots from the core vocabulary, sufficient to justify a conclusion that these two super stocks are genetically related, or are the lexical agreements that exist of a lesser order, to be explained in terms of borrowing/substratum or the like?

The answer he found was that such a corpus does not exist and that such lexical agreements as do occur are to be explained by a relationship in which an unrelated AA linguistic stratum had been superimposed on an AT substratum. Benedict recommended in conclusion that the Austric hypothesis be abandoned.

In 1976, H.L. Shorto defended the Austric hypothesis, citing 52 of some 200 AA/AN lexical comparisons collected, and noting that three fifths of this total correlate to Benedict's AT reconstructions.⁶ He admonished Benedict for undervaluing the hypothesis' strength and understating the extent of the AA/AN core vocabulary agreement, but hesitated to reproach him due to the slow progress of AA historico-comparative studies and general inaccessibility of the AA lexical data. Shorto concluded that he could not accept Benedict's inference that AA and AN, the AT group most likely to have provided the proposed substratum, are genetically unrelated or follow him in discarding the Austric hypothesis.

In 1977, Gérard Diffloth also took issue with many of Benedict's conclusions about the AA/AT relationship. He rejected certain proto-phonemes set up by Benedict for AA and one of its families, proposed others in their stead, and showed that the typological differences between PAT and EAA, which the AT substratum would most likely have influenced due to geographic reasons, are much greater than Benedict had claimed. He argued that if his proposals are correct, then the evidence would actually contradict the hypothesis that a substratumized AT group had ever existed, and the question of an AA/AT genetic relationship would remain completely open. Diffloth also asserted that the possibility of a direct, but very ancient historical grouping cannot be dismissed until our knowledge of AA has been greatly expanded.

More recently (1989), Diffloth has again raised the question of the placement of AA and expressed annoyance at Benedict's dismissal of the Austric hypothesis without any precise appraisal of Schmidt's 215 potential cognates, noting that Pou, Jenner, and Shorto had introduced in the above referenced articles dozens of new possible cognates. He commented that while Benedict seemed to be looking for a case where the majority of the "basic" words had not been replaced or semantically

⁶ I would like to thank Helen Cordell of the School of Oriental and African Studies, University of London, for providing me in February 1990 a copy of Shorto's paper. Benedict's response to this paper has not been available to me.

altered, those words are subject to change and replacement in many sociolinguistic situations and cultural contexts. Diffloth also concluded that "Benedict may detest it, but the Austric hypothesis is still very much alive."

In essence, these discussions with their diametrically opposed conclusions represented little more than another inconclusive round in a debate that has now lasted nearly nine decades. But the general impression they made was discouragingly pessimistic. The very few correspondences discerned by Benedict in the core vocabulary increased immensely the gravity of his doubts about the reality of Austric. And while Shorto's and Diffloth's findings did much to counterbalance Benedict's skepticism, their assessment that a genetic linkage is still possible also presaged a continuing wait of many years, if not decades, before the Austric question could be finally and conclusively settled.

1.3 New Evidence for Austric: A Serendipitous Discovery

After reading the papers of Pou and Jenner, Benedict, and Diffloth in 1977 and 1978, I for one was not encouraged to delve further into the matter or even acquire Benedict's book. However, when I finally bought this volume in 1983 to familiarize myself with the AT stock, I was pleasantly astonished to perceive in it a totally unexpected abundance of lexical matchings between Benedict's AT reconstructions and AA etyma in the Vietnamese etymological files I had been compiling off and on since 1975. Within a few days, I had some 200 sets of possible cognates jotted down, and it was rapidly becoming clear that Benedict had missed something in his evaluation of the Austric hypothesis. Even more amazing was the burgeoning impression that his AT reconstructions might embody not only the means to uncovering the missing lexical proof of the Austric superstock's existence, but might also provide the heuristic key needed to unlock the enigmatic past of AA itself.

What began then as a serendipitous discovery has since become a nine-year long, off-and-on again trek on the track of Austric, which thus far can be divided into three phases. During the first, progress was extremely slow, but with the aid of Benedict's AT reconstructions and a variety of heuristic techniques, small breakthroughs were repeatedly made, with the result that my comparative lexicon increased to some 400 sets of possible AA/AT cognates. A study using them all was written in late 1983, but never finalized because other events led in me in an opposite direction.⁷

The second phase began in late 1988 when the first draft of this paper was begun.⁸ During its preparation, the general outline of the PAA and Proto-Austric

⁷ It is unlikely that I would have continued this research, had not a bizarre twist of fate turned me back to it. After going into business in early 1984, time for my linguistic interests had waned to nothing long before person(s) unknown placed plastic explosive underneath our business van on March 9, 1987. My survival of this incident, which left me partially disabled, was so close to being a miracle as to make me wonder for what purpose my life had been spared. Thus far, I have found none better than this research, and its resumption and continuation have helped me survive and overcome a most difficult time.

⁸ This paper was completed in March 1990, revised in May 1991, and completely rewritten in August 1992. My thanks to the Mon-Khmer Studies editorial staff and especially David Thomas

phonological and morphological systems was worked out, many new details of the AA diachronic evolution were discovered, and the cognate count rose to 640, a little more than Benedict's 1975 total supporting the AT hypothesis.

Until 1990, my comparison was focussed only on the AT data in Benedict 1975. The third phase began with acquisition of new AN material (primarily Wurm and Wilson 1983, but also Blust 1970, 1980, 1983–4), and again I was astonished, for there was such an abundance of comparable data to be seen that it seemed as if no one, including myself, had ever really looked before at the lexical correspondence existing between AA and AN. Since then, I have stopped counting, but the new lexical evidence for Austric probably now amounts to some 1,000–plus cognate sets, including 100 for anatomical terms, 140 for fauna names, etc., with much of the new AN material yet to be fully explored.

2. Probatory Problems for the Austric Hypothesis

2.1 *Problems Galore: Which Is Central?*

2.1.1 *General*

The cursory review of previous attempts to verify the Austric hypothesis presented in Sec. 1.2 makes clear that they were unsuccessful because one general problem, finding really convincing evidence, could not be overcome. In the proposed Austric grouping, where the time depth is several millenia and a multi-stage hierarchy of subgroupings colligates some 870 languages, one could expect to encounter an enormous number of diverse comparative problems. It is not necessary, of course, that all be solved before the Austric hypothesis can be affirmed. But if this study is to succeed where the others failed, it is essential that the problems which are crucially probatory for the existence of Austric be precisely identified and insofar as possible systematically solved.

2.1.2 *Specific*

2.1.2.1 *Theoretical*

All languages change over time, and two general types of language change can be distinguished. Internal change occurs whenever a language which has not been introduced from somewhere else spontaneously replaces another in an area and/or community, while external change occurs when the replacement is made by a language imported from outside the area and/or community (Katičić 1970: 37, 63). The languages thus linked by internal change can be regarded as different forms of one changing language; hence, they evidence linguistic descent, and languages connected in any way by descent are said to be genetically related. According to Radoslav Katičić, the only effect of internal change which is positively identifiable is regular sound change, and he calls it the token by which the genetic relatedness of languages can be recognized.

for their patience with me as I wrestled with the problems of how to present an appropriate introduction to an enormously complex and difficult subject.

Therefore, in order to verify the Austric hypothesis, a genetic relationship must be shown to exist between AA and AT. Katičić (1970: 26) generally defines this type of linguistic relationship as partial correspondence due to common origin. Raimo Anttila (1972:319) identifies the general types of correspondence that constitute proof of such relationship as (1) multiple agreement in the basic (core) vocabulary with sound correspondence and (2) considerable and frequent agreement in sound correspondence and grammatical formants.

However, Katičić (1970: 68, 76, 84, 91) shows that only a specific type of linguistic correspondence is characteristic for genetic relationship, and it is the type which appears as the effect of regular sound change. He describes three kinds of such genetic correspondence: (1) *phonological*: Recurrent correspondence of phonemes as to their position in phonemic strings expressing morphemes with correspondent semantic contents; (2) *morphological*: Phonemic correspondence of any part of phonemic strings expressing word forms but only of such parts having the same position with respect to the other parts of the words compared; and (3) *syntactic*: Correspondences of sentences, clauses, or syntagms with phonemically correspondent strings serving as the expression of at least a part of those morphemes which mark syntactic relations.

The genetic phonological correspondences of related languages encode allophones of their parent language (cf. Katičić 1970: 77ff, 106f.). If the complementary distribution of those correspondences can be determined, these allophones can be classified, either singly or as groups of partially like phones,⁹ into phonological units (proto-phonemes) which in turn form phonemic strings expressing morphological, syntactic, and lexical units (proto-forms). The inventories and distribution of these units constitute a historical diasystem which can be interpreted as a description of the proto-language encoded in the correspondences of its descendants. Katičić further defines genetic relationship as the possibility of constructing such a diasystem for a given set of languages. If it is possible to construct one, the languages are genetically related; if not, their genetic relationship cannot be proven.

2.1.2.2 Methodological

The central procedure used to produce evidence of a genetic linguistic relationship is the comparative method. This method has as its principal objectives comparison of the potentially correspondent phonemic strings of two or more languages to determine if genetic phonological correspondence exists between them, to construct the diasystem (proto-language) coded into any such correspondence found, and to explain this correspondence as the result of regular sound change operating on supposed morphs of the proto-language thus established.

In language groups where not all related languages have a common direct ancestor, the method is subsequently applied to a proto-language as developed

⁹ Cf. Hoenigswald 1950: 358: Partially like sets (of sound correspondences) occurring in mutually exclusive environments are taken to be continuations of one and the same phoneme of the protolanguage.

above and its next closest related (proto–) language(s), to reconstruct the next higher level proto–language, and so on. In the case of Austric, since the parent language of the AA languages has not been reconstructed and many gaps exist in the data (see discussion on teleoreconstruction in Sec. 2.3.3 and comments in 3.1), the method must be applied to any sets of potentially correspondent lexical forms found to exist between PAT and any of the modern AA languages.

Linguistic change, whether of the internal or external variety, tends to replace diachronically all elements of a language; hence, its descendant languages consist of sets of elements which are inherited and for that reason correspond partially and those which are not inherited and may or may not correspond. In producing evidence of a genetic relationship, the comparative method would be most effective if the compared elements were all inherited and most of the correspondences were genetic as defined in Sec. 2.1.2.1. Such an ideal situation occurs rarely, if ever, for the structure and lexicon of languages are always affected to some degree by external and irregular internal change. In theory, these effects should be identified and the irregular correspondences and externally replaced elements should be separated from the regular correspondences and inherited elements before comparison is begun. In practice, this is very difficult to accomplish, and Katičić (1970: 122) argues that the language as a whole must be classified before we can know what is inherited and what borrowed. Since few, if any, of the AA and AT languages can be regarded as wholly classified and the lexicon of a great many is barely or not at all known, the comparatist should initially exercise great care in deciding what is borrowed in these languages.¹⁰

2.1.2.3 *Crucial*

The above description of the specific theoretical and methodological problems suggests that there are only two crucially probatory problems for the Austric hypothesis, to determine 1) whether partial correspondence due to common origin exists between AA and AT and 2) whether evidence of such correspondence can be produced through application of the comparative method.

2.1.3 *Central*

On a practical level, the two crucially probatory problems can be viewed as the two sides of a single comparative coin because in each case the solution process must begin with discovery of potentially correspondent lexical forms. Consequently, one might infer a priori that determining whether such lexical correspondence exists is the most basic practical problem for any comparative study. As indicated in Sec. 1.2, this question has, in fact, posed the greatest difficulty for all previous Austric studies; therefore, the central problem in investigating the validity of the Austric hypothesis is discovering such lexical evidence.

¹⁰ As opposed to past practice which has seen many facile conclusions about which language is the loaner and which the borrower. AA and AT have undoubtedly exchanged much linguistic material, but a solid basis on which to decide which borrowed what from which, particularly at the more remote time levels involved in this relationship, remains to be established.

In summation, the Austric hypothesis can be regarded as affirmed when and if lexical correspondences can be accumulated in sufficient number to demonstrate through application of the comparative method that genetic phonological correspondence exists between AA and AT in sufficient quantity to construct a historical diasystem interpretable as the Austric proto-language. The reader should note, however, that since the term “sufficient” cannot be precisely quantified by any theoretical or methodological means, any decision as to whether or not that affirmation has been adequately accomplished must be performed a subjective one.

2.2 *The Lexical Evidence Problem*

2.2.1 *Does an Austric Comparative Lexicon Exist?*

The lexical evidence published heretofore in support of the Austric hypothesis comprises perhaps 350 sets of possible lexical correspondences, cf. Schmidt 1906 (215 sets), Pou and Jenner 1975 (65), Shorto 1976a (52), and smaller collections in Kuiper 1948, Pinnow 1959, Shorto 1961, and Benedict 1975, virtually all comparisons between AA and AN. This published evidence has posed a sort of dilemma for comparatists because—all previous conclusions notwithstanding—the Austric hypothesis really cannot be affirmed or refuted by it.

As Shorto also comments, one might suppose this quantity of data sufficient to establish genetic relationship, but closer examination reveals some significant difficulties. Core vocabulary agreements are conspicuously rare, very few sets meet the semantic identity criterion for rigorous correspondence, and there is a disproportionate amount of semantically specialized, non-basic vocabulary of a type often subject to being borrowed. This “probative” data also clashes qualitatively with the probative grammatical congruences signaled by Schmidt as existing between AA and AN. These shortcomings raise serious doubt that an Austric comparative lexicon really exists.

The generally poor quality of the published lexical evidence could be interpreted as proof of a non-genetic relationship, as Benedict did in proposing an AA/AT-substratum consortium, or an extremely ancient historical grouping, as suggested by Diffloth. In an evaluation of the AT hypothesis, Lawrence A. Reid (1988: 23) observes that:

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.

In this light, one might suspect that no such convincingly large body of lexical forms exists for Austric due to the remoteness of its split into AA and AT.

Shorto and Diffloth both infer a considerable time depth for that split, but at present a more precise assessment can be made only on glottochronological grounds, and indirectly at that. In his new book, *Japanese/Austro-Tai*, Benedict adds the Japanese-Ryukyuan family to AT, placing it with AN in a new AJ

component, and estimates (1990:156) that the Japanese–Ryukyuan and PAN-speaking peoples split off from the PAT-speaking “core” at 5,000 B.C., “give or take a millenium or so.” Although exactly how he came to that date is not revealed, Benedict references the work of Robert A. Blust, according to whom PAN was presumably spoken around 5,000 B.C. (1980:13) and likely split into Formosan and Malayo–Polynesian circa 4,500 B.C. (1988:54). Since the division of AK into AJ and Kadai is antedated in the new AT classification scheme by the split of AT into AK and Miao–Yao, the appearance of AT itself would have to be set back well before 5,000 B.C.—let us say, for argument’s sake, at 6,500 B.C., plus or minus whatever.

On the other side of the Austric house, glottochronological computations based on the cognate percentages determined by Franklin E. Huffman (1978: 8) to exist between certain AA subgroups indicate that PAA had begun to disintegrate into the dialects which would become the modern AA subfamilies by circa 2,000 B.C., give or take 500 years.¹¹

Now, the enormous gap between that date and the tentative dating of PAT at circa 6,500 B.C. (or even of PAN at 5,000 B.C.) raises some formidable questions about this entire chronology and our present assumptions about Austric and AT. Three possibilities come easily to mind: 1) AA and AT are not genetically related, the least likely explanation as this study will show; 2) the renowned frailties of glottochronology are at fault; hence, the AT dates may need deflation or perhaps the AA date is ultra-conservative; and 3) the current classification scheme is in some manner erroneous; hence, Austric (AAplus ?) or AA alone could belong somewhere in the AT family tree, a thought to gladden Benedict’s heart, or Austric could consist of a smaller AT and an other subgroup comprising AA and one or more of the other language families currently assigned to AT. Plausibly, both 2) and 3) apply.

At the present stage of investigation, the emerging AA proto-language is typologically very close to PAN, a classificatory statement which cannot be applied to any other AT subgroup. But the PAA phonological system stands closer to that of PAT and cannot be derived from the PAN phonological system, or vice versa. The PAA morphological system stands apart from that of either PAT or PAN, and the best explanation of their differences currently appears to be that the PAA and PAT systems evolved out of an older common morphological system. On this slim evidence, it will be assumed for the purposes of this study that Austric was indeed the parent language, PAA and PAT its two direct descendants.

It is probably also safe to assume that the interval since the separation of AA and AT amounts to 4,500 years at the minimum, 8,500 as a possible maximum. Although controversial, the glottochronological method tells us that after a 4,500-year separation, one could expect to find 150 cognate sets per 1,000 words of original vocabulary, but only 28 after 8,500 years. These ratios could vary considerably if the actual vocabulary attrition rates in either stock were significantly higher or lower than the 19% per millenium average used in the above calculations.

¹¹ Geoffrey Benjamin (1976: 83) estimates that Proto–Aslian had broken completely away from the other AA languages by before 5,970 Before Present. However, he uses a 14% vocabulary attrition rate instead of the 19% rate used in my computations. Using the latter rate, his date would be 4,270 B.P. or about 2,280 B.C. and within the range cited in the text.

The total number of sets retained is less easily estimated due to such unknowns as the total amount of original vocabulary, but could well be higher than generally presupposed. In this regard, it is note worthy that the Austronesianists now count their PAN reconstructions in the thousands, cf. Blust 1980: 6.¹²

In summation, one may draw three tentative conclusions on the basis of the preceding analysis. First, the time depth is not great enough for the lexical elements common to the earliest stage to have been reduced to a negligible quantity; therefore, an Austric comparative lexicon does exist if AA and AT are genetically related. Second, the total number of lexical correspondences in that lexicon is indeterminate, but could be larger than generally thought. For example, if the original Austric lexicon contained only 5,000 items, perhaps a low figure in view of the AN findings, the extant comparative lexicon could comprise 140 to 750 cognate sets, depending on the actual time depth and attrition rates, generally more than Schmidt's 215 sets. Third, these cognates should denote mostly basic vocabulary, and the probative lexical evidence recovered, even if meager, would consequently be quite convincing.

2.2.2 *Discovering the Discovery Problem*

In 1983, the easy discovery of some 200 sets of possible AA/AT cognates initially promised quick resolution of the lexical evidence problem, but my search soon ran into the same invisible but nevertheless impenetrable wall that previous searchers had apparently encountered. At the same time, two observations suggested that additional lexical evidence might be discoverable. First, a huge number of "possibles," such as AT *q[a]lu[b](u) and Santali bɔhɔ 'head', evinced a certain vague aura of comparability, but had to be rejected as non-corresponding due to various phonetic and/or semantic irregularities. Second, it appeared that in a large number of cases AA might have undergone all of the more exotic transformations Benedict had discerned in AT, i.e. stress shift, stem splitting, cluster mutation, and vowel transfer, and these might account for many of the unexplainable irregularities. These observations also suggested that maybe the heuristic difficulties were not the result of looking for what is not there (non-existent cognates destroyed by diachronic change), but of not seeing what is there (cryptocognates disguised by such change).

The suspicion that I was looking at an Austric lexical forest but not seeing its cognate trees was slowly substantiated as some of the cryptocognates were tentatively deciphered. In the dubious 'head' set, for example, it was seen that AA and AT probably manifest formal correspondence and similar development, i.e. a common root, *q[o]b[u], had been infixed in AT to yield *q[a]lu[b](u), whence AN *qulu[h] by vocalic assimilation and canonical reduction, but had evolved in AA to Santali bɔhɔ, either by metathesis or reduction of a reduplicate. The additional lexical evidence accumulated in this fashion was not large at first, but it clearly indicated that AA and AT might be linked by lexical correspondence and parallelism in historical development of a magnitude not previously detected or

¹² Laurent Sagart (1990: 3) points out that the large number of AN reconstructions may be misleading because different items commonly share the same "root". The morphological analysis presented in this paper (Sec. 3.4 and 4.3) also suggests that many of the numerous AN proto-forms are allomorphs (derivatives) of a much smaller number of lexical morphs (stems).

suspected, and this linkage might be concealed by phonological and semantic divergence of such nature that strict application of the comparative method could not penetrate the obscurity to reveal the true character of the AA/AT relationship.

The above-described impressions were not easy to accept at first because they clashed sharply with many of the traditional views on the nature of the AA/AT relationship and the early structure and diachronic evolution of AA. But they also implied two important conclusions which could not be ignored for long. First, the lexical evidence problem is not a theoretical issue resulting from absence of genetic correspondence between AA and AT, but a methodological matter confirming the existence of a problem in the discovery of such evidence. Second, probative lexical evidence must exist in larger quantity than heretofore found, and it should be discoverable if more appropriate and efficient comparative procedures could be developed.

2.2.3 *The Search for a Solution*

In the process of coming to those conclusions, it slowly became clear that before a formal comparison of AA and AT (or indeed of the AA languages themselves) could be begun, one must learn first how to compare them. The immediate task was, however, to assemble a quantity of comparable lexical sets which hopefully would be sufficient to permit that learning process to take place.

Between 1983 and 1990, a variety of heuristic techniques were developed and tested, and each served to add additional sets of potentially correspondent lexical forms to the Austric comparative lexicon and to reveal additional details about the manner in which AA and AT correspond. The most successful of these techniques was based on a “rule” proposed by Benedict in 1939: “One must use dubious semantic comparisons to establish phonetic equivalences, and vice versa” (Shafer 1966:10).¹³ It helped add several hundred sets to the lexicon.

This rule seems to sanction relaxation of the theoretical and methodological requirements given in Sec. 2.1.2, and Benedict (among others) has cautioned that its misuse could lead to disastrous results. In practice, the relaxation should be only temporary and exploratory. The AA/AT comparative lexicon is characterized by phonological and semantic equivalences which initially appear to be irregular and sporadic, but can often be shown to be regular and recurrent upon further study. Benedict’s rule at least keeps them under scrutiny long enough to determine if corroborative examples can be found or the conditioning factors can be identified. In the case of semantic correspondence, Katičić and others have admitted that it is only a probabilistic concept; its presence adds credibility to the probative evidence of phonological correspondence, but its absence need not totally invalidate that evidence. As it happens, semantic divergence is a typological trait of the AA/AT genetic relationship.

Since 1990, the quantity of potentially correspondent lexical sets which can be set up between AA and AN has made the explorative approach described above largely irrelevant and raised the possibility that a conventional application of the

¹³ The quote is from Benedict’s “Semantic Differentiation in Indo–Chinese”, *HJAS* 4, 1939, pp. 213–29.

comparative method can be undertaken for AA and AT, whenever the internal reconstruction of AA can be accomplished (cf. Sec. 3.1).

In summation, the cause of the Austric lexical evidence problem has never been an absence of comparable data, but a lack of insight to the historical dynamics of AA and their role in concealing lexical and sound correspondence (cf. Sec. 3.2). Thanks to Benedict, whose AT reconstruction has so often served as a guiding light, the insight necessary to discovery of the needed probative lexical evidence can now be developed, and it appears that the discoverable evidence will be more than sufficient to demonstrate the partial correspondence between AA and AT required to verify the Austric hypothesis.

2.3. *The Megalocontroversy*

2.3.1. *The Theory of Large-Scale Comparisons*

In a recent disquisition on very large-scale linguistic comparisons, James A. Matisoff discusses the megalocomparisons performed by Joseph H. Greenberg on Amerindian and Benedict on AT, critically appraises their megalocomparative methodology, and mentions making in earlier papers a distinction between three sorts of linguistic comparison (and observing that “different species of maniacs” are attracted to each sort).¹⁴ In a footnote (1990: 108), Matisoff defines this theory of large-scale linguistic comparison as follows:

Very roughly speaking, MICROCOMPARISON can be practiced on close-knit families like Romance, Loloish, or Tai, with a time depth of not more than about 2000 years. MACROCOMPARISON is appropriate for farflung but demonstrably valid groupings like Indo-European or Sino-Tibetan, with time-depths of up to about 6000 years. MEGALOCOMPARISON takes on any more remote relationships, where sound-correspondences are not regular and putative cognates are few, so that chance rivals genetic relationship as the explanation for perceived similarities.

2.3.2. *Austric: Megalo- or Macrogrouping?*

Since Matisoff has preemptively classified AT as a megalogrouping, any comparative study of AA and AT must be regarded ab initio as a megalocomparison. This classification automatically ascribes to the AA/AT relationship the attributes of Matisoff’s definition, but if this relationship is well described by those attributes, it cannot possibly be genetic. Consequently, this classification poses an additional probatory problem for the Austric hypothesis, which apparently must be solved by proving that Austric is not a megalogrouping.

¹⁴ The referenced papers by Matisoff are “Austro-Thai and Sino-Tibetan: An Examination of Body-Part Contact Relationships”, *Genetic Relationship, Diffusion, and Typological Similarities of East and Southeast Asian Languages*, edited by Mantaro J. Hashimoto, Japan Society for the Promotion of Science, Tokyo, 1976, pp. 256–89, and “Problems and Progress in Lolo-Burmese: Quo Vadimus?”, *LTBA* 4.2.

In Sec. 2.2.1, it was shown that the time depth between AA and AT could vary any where from 4,500 to 8,500 years (or even more). Since Matisoff's 6,000-year parameter falls near the middle of that span, there probably is no cogent reason why Austric could not be reclassified timewise as a macrogrouping, even if the actual time depth lies on the high side of that boundary.

It will become evident in the course of this series that regular, recurrent sound correspondence is easily demonstrated between AA and AT and that a large number of credible lexical correspondences (cognates), including many in the core vocabulary, can be found.

In conclusion, the demonstrable evidence clearly supports genetic relationship, rather than chance, as a better explanation of the similarities between AA and AT. It also suggests that Austric is better classified as a macrogrouping. In his discussion, Matisoff seems to have overlooked the possibility that regular correspondence and abundant cognation might be found where formerly little or none was perceived, thereby requiring such reclassification or perhaps a new term for demonstrably valid megalogroupings.

2.3.3 *In Defense of Benedict's Megaloprocedures*

In his AT studies, Benedict adheres to the comparative method insofar as possible, which is also the methodological approach taken in this study (despite the forays into heuristic adventurism). Greenberg took a quite different course, and his methods need not concern us here. But Benedict also developed some methodological innovations which Matisoff finds objectionable and describes as follows (1990: 116f.: 1) *teleoreconstruction* (Benedict's term): "The method of leaping back to the level of the proto-language, without being deterred by all the detailed problems which arise in the individual sub-groups;" 2) *pseudomicritizing devices*: "Notational attempts to make the speculative seem rigorous" and ensure the "apparent regularity of correspondence," which include the next three; 3) *proto-form stuffing*: Reconstruction of "proto-forms that are so complex canonically that no given combination of proto-entities is likely to recur very often, thus obviating counterexamples;" 4) using "a rich apparatus of *bracketings* to optionalize virtually every portion of the proto-form;" and 5) *split cognates* (Benedict's term): "Cognates that have reflexes of at most one given proto-phoneme in common, since they descend from different syllables of a polysyllabic etymon."

Although probably meant to be humorous, Matisoff's descriptions potentially impugn the integrity of the serious megalocomparatist (doubtlessly an oxymoron for Matisoff), who in using such procedures may not really be trying to evade the rigors of the traditional methodology or "doctor" his evidence. In this study, teleoreconstruction, bracketing, split cognates, and some equally suspect heuristic techniques like Benedict's "rule" have been applied with no little measure of success; pseudomicritizing devices, such as proto-form stuffing, have not been used.

In an ideal linguistic world, the comparatist starts at the lowest level in each subgroup and works upwards, level by level, taking care of all those detailed problems at each step. In the real world, data may not be available from all dialects

of one language in the first subgroup or from all languages in the next most closely related subgroup. Or data is not available from this subgroup but is from a third less closely related subgroup. In such cases, the comparatist has only two options, wait until the missing data become available or use teleoreconstruction to bridge the data gaps.

In AA, where the total number of languages and major dialects cannot yet be precisely enumerated, relatively few descriptive studies and dictionaries are available, and few historico-comparative studies exist for any level, the choice seems obvious. In such circumstances, teleoreconstruction could be regarded as a necessary evil, but it also has some redeeming virtues. The studies of Bernhard Karlgren, Otto Dempwolff, and Benedict on Ancient Chinese, AN, and AT, respectively, all teleoreconstructive to some degree, have stimulated considerable progress in Sino-Tibetan, AN, and Tai-Kadai linguistics. If this study produces nothing more than a similar effect in AA, the effort will be more than well compensated. It also cannot be less advantageous to attack the comparative problems of AA from two levels rather than just one alone.

As for Benedict's other new procedures used here, it should become clear in the course of this study that they are not only perfectly appropriate to the situation, but represent in fact a brilliant insight by their originator to the structure of the earliest stages of AA and AT and the historical changes which have affected it.¹⁵ Although still dim and imprecise, the emerging view of PAA (and its Austric ancestor) is one of a highly synthetic, perhaps even agglutinative, language in which mono- and disyllabic lexical morphs combined with grammatical morphs to form canonically complex polysyllabic words. This structure was subsequently altered in two major ways. In one, affixal morphs were incorporated into stems, and these stems often recombined with the same or other morphs to produce additional strata of canonically complex lexical forms. In the other, a perhaps originally variable stress accent shifted in various patterns, and unstressed syllables mutated or dropped off, producing the split or truncated stems now evidenced as split cognates. EAA and mainland AT have actually undergone two such shifts.

These developments can be briefly demonstrated by the 'red/dark-colored/redden/shame(d)' etymology cited by Matisoff (1990:116), cf. also Benedict 1975: 361 and Sec. 3.4 below.

PAT	PAA	Modern AA
*i γ a(?)i	*i γ a[q]i	Sora (P396) <i>jγe:</i> 'red'
*i γ a η	*i γ a/n/[q]/a/t	Mundari <i>jeηe'</i> 'deep redness, crimson'
*(q)b/i γ a η	*b/i γ a/n/[q]	Bahnar <i>brηeη</i> 'rust'

In conclusion, Benedict's methodological innovations can be regarded as adaptations to the specific requirements of the languages and comparative tasks at hand. As such, they need pose no theoretical challenges to orthodox methodology. These procedures are clearly useful in Austric, and this is probably true because AA and AT are genetically related; they may not be practical for other language groups.

¹⁵ I still subscribe to this evaluation of Benedict's methodology, even though he has since (apparently) abandoned some of his innovations, such as "bracketing" (usage of slashes to denote morphemic boundaries, to be precise). To me, they still seem to be rather useful.

3. The Reconstruction of PAA

3.1 *Probatory Problems for the AA Hypothesis*

Over the past 28 years since Thomas surveyed the state of affairs in AA and MK comparative studies in the inaugural volume of this journal (1964:149–63), much progress has been made in expanding our knowledge of these languages. Unfortunately, little of it has served to provide anything like the bridge work which appears to be needed in order to span the enormous gap that yawns between the generally low-level reconstruction accomplished thus far and the AA proto-language as it appears to be reconstructible on the basis of the present AA/AT comparison. The excellent studies of Heinz-Juergen Pinnow (the last published in 1966) are the sole exception, and his AA reconstruction now seems excessively conservative. Moreover, some of the problems deplored by Thomas in 1964 still remain, in particular the general paucity and inaccessibility of AA lexical data.

The absence of a formal PAA reconstruction seriously hampers any effort to compare rigorously AA and AT, and the internal comparison prerequisite to that reconstruction poses in and of itself a challenge equally great to that posed by the probatory problems of the Austric hypothesis. As in the case of the latter, the central investigative problem in AA is also discovery of correspondent lexical evidence, but here the difference is one of quality, not quantity. EAA (Aslian, Khasi, MK, Palaungic) is well represented in the published lexical material to which the writer has had access, but Munda (Western AA) and Nicobar (Southern AA) are not. Munda and EAA are typologically quite different and an important question in AA comparative studies has always been which branch is representative of the parent language. For this reason (among many others), any attempt to reconstruct PAA must take the Munda evidence into heavy consideration. Since the writer's Munda material is especially lacking in core vocabulary, the kind in which inherited elements and genetic correspondence are most likely to be found, the desired formal PAA reconstruction can be undertaken only in a limited way, and much of the reconstruction presented here as PAA may really be valid only for Proto-EAA.

As a result of the current state of affairs in AA linguistics, this study has been obliged to take on two functions, production of evidence supporting the Austric hypothesis and production of evidence supporting the AA hypothesis. In this section, the focus is on the latter type, but it is supported by the former type. The process by which our conclusions about the nature of PAA are reached is thus a sort of inverted reconstruction, an accepted adjunct procedure of the comparative method. But since this procedure is typically applied after a proto-language has been reconstructed on the basis of its descendant languages, those conclusions are subject to the same criticisms leveled at teleoreconstruction (Sec. 2.3.2). Hence, the AA proto-units presented here can be offered only as tentative (i.e. work-in-progress) reconstructions.

3.2 *Some New Aspects of AA Linguistic History*

3.2.1 *General Comments*

In Sec. 2.2, the Austric lexical evidence problem was attributed to lack of insight into the historical dynamics of AA and their role in concealing lexical, phonological, and morphological correspondence, both within AA and between AA and AT. In the writer's opinion, the same lack is at fault in the generally slow rate of progress in AA comparative studies. In this subsection, those claims will be substantiated by examples of the results of certain processes of diachronic change which have drastically altered the face of PAA, thereby creating the apparent gap and the various difficulties which comparatists have encountered in attempting to verify the Austric hypothesis and launch AA comparative studies towards a successful conclusion. These processes, stress shift, cluster mutation, vocalic transfer, and affix incorporation, have been mentioned before in one way or another in the literature (cf. Pinnow 1959:18, 23f., 439f., Shorto 1976:97f., Diffloth 1989:6), but with the exception of stress shift, their importance has not been adequately appreciated or understood.

3.2.2 *Stress Shift*

Benedict (1975:153) proposes that PAT had a variable stress accent which became fixed in the AT dialects. Shifting and fixation of this stress produced different patterns of canonical reduction which he (1990:20ff.) terms left, center, and right according to whether an initial, medial, or final syllable, respectively, was lost. When stress was fixed on different syllables in the same or different language(s), different reduction patterns resulted and stem splitting occurred, as in AT ***mapra*, PMY **may C*, PKD **praa A* 'eye'.¹⁶

In AA, fixation of a once variable stress accent is also indicated, and the related phenomena of stress shift, canonical reduction, and stem splitting have also occurred. The mechanics of these change processes are not yet well understood because the situation is complicated by the fact that two distinct phases of stress shift can be distinguished. During the first, there seems to have been a tendency for stress to be shifted progressively (i.e. to or toward the initial syllable), as in Miao-Yao, and for canonical reduction center and/or right to follow, e.g. **bu[q]aya > *biya > Khmer *kraboe* 'crocodile'.¹⁷ The effects of this shift have been obscured by those of the second phase during which stress was unidirectionally focused on the final syllable, and the initial syllable was dropped (canonical reduction left) or compressed into the final (often accompanied by vocalic transfer), e.g. **biya > Katu *bayua* [biyɨə] 'crocodile'. The second phase—also known as the trend to monosyllabism—has affected only EAA (and mainland AT).**

¹⁶ The double asterisks denote AT proto-forms found in Benedict 1990.

¹⁷ Except where copied directly from sources cited in the reference section, transcriptions of Khmer words are based on Jacob 1974:16–17; any errors therein are my sole responsibility.

Set 1. Stress Shift.¹⁸

AT	PAA	AA
*[(m)b]uqay(ɿa), AN *buqaya[ʔ,h] ‘crocodile’	*bu[q]aya	Khmer <i>kraboe</i> , Katu <i>bayua</i> ‘crocodile’, Chrau <i>raya</i> ‘spotted monitor’
**kaʔ, AN *kaʔen ‘eat’	*kaʔən	Pacóh <i>pancan</i> [pɔnka:n], Jeh <i>kokǎ</i> [kɔkaʔ] ‘chew’, Thavung <i>ʔan</i> ‘eat’
AK *ʔa(n)ji ‘sibling (older)’, AN *sua(n)ji ‘younger sibling’	*(i,a)(n)ci(q)	Kharia (P38) ¹⁹ <i>aji</i> ‘elder brother’s wife’, VN <i>anh</i> ‘elder brother’, Pacóh <i>achai</i> ‘older sibling of the same sex’
*r[a]dzay, PMY *ray ‘sharp’	*rajay	Pacóh <i>raiq</i> ‘sharpen to a point’, Jeh <i>pla jei</i> ‘sharp’
AN *qaRem ‘(scaly) anteater’	*(i)(n)ʔ[ə]ma	VN (*n’jim >) <i>dím</i> , <i>nhím</i> , Jeh <i>jima</i> , Khmer <i>pramā</i> ‘porcupine’

3.2.3. Cluster Mutation

Benedict (1975:175) shows that in the early AT dialects, certain consonant clusters coalesced into single phonemes. In AA, the same type of phenomenon has taken place (Set 2), but thus far the available examples are too few to tell much about the conditions under which it occurred.

Set 2. Cluster Mutation

AT	PAA	Modern AA
*[qilqɿ][uʔ]. AN *t]elur ‘egg’	*[q]liʔu[ʔ]	Juang <i>susuterɔ</i> , Pacóh <i>tireal–tirǎl</i> ‘egg’, Thavung <i>iluul</i> ‘lay an egg’
*[(m)ba]ntriŋ ‘cattle, buffalo’, AN *baNTiŋ ‘wild cattle’	*(b)[l.r]iŋ	Sora (P263) <i>taŋ–li:n</i> ‘cow, cattle’, PM * <i>priəŋ</i> ‘water buffalo’, * <i>kndiŋ</i> ‘gaur’
*(m)b[ə]ʔyuk, AN *be[r]juk ‘monkey’	*ta(m)bəʔyuk	Brou <i>tamúr</i> , Pacóh <i>rboq</i> , Sengoi <i>dok</i> ‘monkey’
*(m)praŋ ‘bee’, Malay <i>indong</i> ‘honey–comb’	*(m)praŋ or *(p,b)raŋ	Sora <i>əðəŋ</i> ‘honeycomb’, Pacóh <i>itang</i> ‘tiny black bee’, Chrau <i>sidang</i> ‘sugar’
*(N)qraŋ[a], AN *talinaŋ ‘ear’	*[qə](n)[qr]iŋ	Kensiu <i>ʔəntiŋ</i> ‘ear’, Rengao <i>ik kring</i> ‘ear wax’

¹⁸ Unless otherwise noted and except for such retrancriptions as ‘b, etc., for implosives, all lexical forms are reproduced with the orthography used in the source documents. In proto–forms, parentheses () denote optionality, brackets [] uncertainty, commas separate alternate forms, and slashes / mark morphemic boundaries. Slashes may also identify isolated affixes, alternate forms, or phonemic representation. Brackets may also denote phonetic representation. Cited AN proto–forms come from a variety of sources, and in some cases I have changed them to reflect the usage of parentheses and brackets in this paper.

¹⁹ Pinnow’s lexical sets are referenced by P and his item number, with K preceding the number and identifying sets found in the consonant section.

3.2.4 Vocalic Transfer

Benedict (1975:183) shows that Kadai and, to a lesser degree, Miao–Yao exhibit a complex system of vocalic transformations, which he labels vocalic transfer. He describes this process as an aspect of the trend to monosyllabism involving the transfer of the vowel in the presyllable to the main syllable. He also suggests that the product of this process is dependent primarily on the nature of the transferred vowel and secondarily on the consonants and stress factors involved.

In AA, the mechanics of vocalic transfer remain little understood, but evidence exists that vowels were transferred progressively and regressively.

Set 3. Vocalic Transfer.

AT	PAA	AA
* <i>tsu[ra]k</i> ‘angle, elbow’, AN * <i>seruk</i> ‘angle, corner’	* <i>cu(r)ʔak</i>	PW * <i>sʔwok</i> ‘elbow’, Nyah Kur (Tha Pong) <i>cəɾwaak</i> ‘fork’, Pacóh <i>choaq</i> [coʔ] ‘use feet to measure’
*(<i>q</i>)(<i>n</i>) <i>tulaq</i> , AN * <i>[t,T]ulaj</i> ‘bone’ AN * <i>lasu</i> ‘genitals’	* <i>c(i,a,u)(n)qaj</i> * <i>lac[u]</i>	Khasi (Amwi) <i>šiang</i> , Lamet <i>cəŋʔaaŋ</i> , VN (* <i>cihaŋ</i> >) <i>xuong</i> ‘bone’ Kurku (PK289) <i>laj</i> , Santali <i>lɔʔj</i> , Ruc <i>luc</i> ‘penis’
AN * <i>dilat</i> ‘lick, lap’	* <i>[d]ilat and</i> * <i>[(i,a)(m)pat</i>	Khasi <i>thyliet</i> [tʰəlet], Bahnar <i>ropiët</i> ‘tongue’, Pearic <i>pat</i> ‘lick’ ²⁰
AN * <i>Ruqaj</i> ‘throat’	* <i>Ru(n)qaj</i>	VN <i>həŋg</i> , Ruc <i>kəŋzəŋ</i> ‘throat’, Sedang <i>rokong</i> ‘mouth’

3.2.5 Affix Incorporation

In Japanese/Austro–Tai (1990:123ff.), Benedict has begun to perceive the importance of affix incorporation in AT. During the preparation of this paper, it became clear that this process has played a crucial role in the evolution of AA and AT. Although it has long been recognized, more or less tacitly, that affixes have been incorporated into stems in AA, the number of languages and affixes involved, the existence of multiple layers of accreted affixes, and the processes’ role in concealing correspondence are only now becoming clear. The ‘first–person singular personal pronoun’ and the numeral ‘one’ present clear examples of such incorporation.

²⁰ Pearic refers to the lexical items cited by Robert K. Headley Jr. in his article; they actually come from several languages of the Pearic branch.

Set 4. Affix Incorporation.

PAT	PAA	Modern AA
**ʔa(ŋ)ku, AN *ʔiaku, *(ʔu)aku, *ʔa(ŋ)kən '1'	*(m)(i,a)(n) qu(a)(n)	*m/i/n/qu (Pareng <i>mingu</i>), *i/n/qu (Kharia <i>iŋ, iñ</i>), *n/qu/a (Khasi <i>nga</i> , VN <i>qua</i>), *a/qu (Riang <i>ʔoʔ</i> , Jeh <i>au</i>), *n/qu (Katu <i>ku</i>), *n/qu/a/n (Thavung <i>kan</i>), *a/qu > *i/aqu (Temiar <i>yɛʔ</i>) *i/qu/a (Nicobar <i>cūə</i>)
AK *ʔitsa, AN *esa, *isa 'one'	*(m)(u,i,a)(n)ca	*i/ca (Khasi <i>shi</i>), *a/ca (Katu <i>chakui</i> 'one person'), *mu/aca (Muong Khen <i>məh</i>), *m/i/n/ca(ʔ) (Mundari <i>miya'd</i> , Bahnar <i>mính</i>), *mu/anca(ʔ) (Kharia <i>mɔyɔ'd</i> , PM *muay), *mu/inca (Juang <i>muiñ</i> , Kotua <i>muiq</i>), *bu/anca (Sora <i>boi</i>)

In the 'one' set, layers of accreted affixes seem to be indicated, as in *ca > *a/ca > *u/aca > *m/uaca and *a/n/ca > *u/anca > *m/uanca, perhaps also *[ʔ]/(i,a)/(n)ca > *m/u/[ʔ]i(n)ca and *m/u/[ʔ]a(n)ca.

3.3 The PAA Phonological System

3.3.1 Methodology

Evidence of genetic phonological correspondence is typically presented through means of lexical sets reflecting all positional and allophonic occurrences of all of the proposed photo–phonemes. To conserve space in this introductory study, the exemplary sets shown below depict only one such positional occurrence. Munda forms are included in these sets whenever possible.

If the requirement of correspondent semantic content were rigorously applied, a great many of the available cognate sets would be unusable. Nevertheless, some AA language usually does have the same meaning as AN or AT, even if the remainder do not. In general, such differentiation has a positive value in that its presence makes it less likely that the lexical correspondence is a result of borrowing and more likely evidence of common origin.

3.3.2 The PAA Vowel System

Thomas (1964:160f.) identified the complexity of vowel shifting as one of the major difficulties encountered in MK comparative studies, and evidently this is no less a problem in Munda, for Pinnow (1959:195) did not even attempt to set up a PAA vowel system as he did for consonants. He did establish */i, ɛ, i, ə, a, u, ɔ/ as the Proto–Munda system, which differs somewhat from the PMK vowels proposed by Shorto (1976:1043), */i, e, ə, a, u, o, ɔ/. Shorto also reconstructed a length feature and three diphthongs, */iə, uə, ai/, while Pinnow felt that neither could be firmly established at the Proto–Munda level.

At present, it is clear that PAA probably possessed at least the same six-vowel system, **/i, e, ə, a, u, o/*, which Benedict (1975:172, 1990:34) reconstructs for PAT. It is quite possible that PAA also had three additional vowels, **/ɛ, ɪ, ɔ/*, at least subphonemically. Certain correspondences suggest this and that they had developed primarily through vocalic assimilation. PAA does not seem to have had true diphthongs.

Set 5. PAA Vowels.

Modern AA	PAA	AT
*i Kharia (P38) <i>aji</i> 'elder brother's wife', Khasi <i>kynsi</i> 'spouse's sibling', PM <i>*kmcii?</i> 'younger sibling'	* <i>(a)(n)ci(q)</i>	AK <i>*ʔa(ŋ)ji</i> 'sibling (older)', AN <i>*sua(ŋ)ji</i> 'younger sibling'
*e Kharia (P305) <i>hepa' d</i> 'near', VN <i>hɛp</i> 'narrow', Chrau <i>gêp</i> 'close'	* <i>(n)Gep(at)</i>	* <i>(e)(N)Gep</i> , PMY <i>*NGep</i> 'narrow'
*ə Santali (P365) <i>jan̄ga</i> , PW <i>*joŋ</i> , PM <i>*juŋ</i> 'foot, leg'	* <i>(n)jəŋ</i>	* <i>() dzəŋ(dzəŋ)</i> , AN <i>*zɛzɛŋ</i> 'stand'
*a Kharia <i>tataŋ</i> , PW <i>*taʔ</i> , Pearic <i>tai</i> 'grandfather'	* <i>(ta)ta(q)</i>	* <i>tata</i> , POC <i>*tata</i> 'father, grandfather'
*u Kharia <i>ma mu</i> 'mother's brother', VN <i>m̄u</i> 'old woman', Mon <i>m u</i> 'parent's younger brother'	* <i>(mu)mu</i>	* <i>[(t)a]mu</i> 'grandparent, grandchild', FO <i>*mumu</i> 'grandfather'
*o Mundari (P311) <i>hɔɔ</i> 'a man, a Munda', Khasi <i>kur</i> 'a clan', Old Mon <i>kulo</i> 'kinsman'	* <i>(n)quʔo</i>	* <i>() (N)[Go]ʔon</i> , AN <i>*hulun</i> 'man (human being)'

Thus far, the AA/AT comparison has been more useful in showing us how the great vocalic diversity came about than in revealing a complete picture of the ancient vowel system underlying that complexity. In the rare cases where the PAA phoneme appears to have escaped most modification, as in most of the above examples, the modern reflexes seem to correlate to the six-vowel system and are typically short and monophthongal. In most cases, however, much shifting has taken place, and the processes responsible for much of it can be identified as stress shift, canonical reduction, vocalic transfer, vocalic assimilation, and various sorts of environmental conditioning. The length feature and most of the diphthongs are clearly products of vocalic transfer and intervocalic loss of consonants; vocalic assimilation has stimulated the creation of new vowel phonemes, such as **/ɛ, ɪ, ɔ/*.

3.3.3 The PAA Consonant System

The PAA consonant system proposed by Pinnow (1959:427), **/p, t, c, k, b, d, j, g, l, r, w, y, m, n, ŋ/*, is far too conservative. The AA/AT comparison indicates that at least 10 additional proto-consonants must be set up, as in the following data table. **[N]* was probably an allophone of the other nasals before the postvelar stops.

Set 6. The PAA Consonants.

*p	t	c	k	q	ʔ
b	d	j	g	G	
	s		x		h
	z		Y		
	l		ʃ		
	r		R		
w		y			
m	n	ñ	ŋ	[N]	/

AA

PAA

AT

*p Kharia (P305) <i>hepa' d</i> 'near', VN <i>hep</i> 'narrow', Chrau <i>gêp</i> 'close'	*(n)Gep(at)	*(e)(N)Gep, PMY *NGep 'narrow'
*t Kharia <i>tataŋ</i> , PW * <i>taʔ</i> , Pearic <i>tai</i> 'grandfather'	*(ta)ta(q)	* <i>tata</i> , POC * <i>tata</i> 'father, grandfather'
*c Santali (PK289) <i>lɔ'j</i> , Ruc <i>lúc</i> , Nicobar (Car) <i>ku lɔich</i> 'penis'	*lac[u]	AN * <i>lasu</i> 'genitalø'
*k Kharia (P216) <i>sɔʔlui</i> , Mal <i>nsook</i> , Ruc <i>usúk</i> 'hair'	*(n)suk	* <i>busuk</i> , AN * <i>buhuk</i> 'hair'
*q Santali (P206) <i>bɔhɔ</i> , Chrau <i>vôq</i> 'head', Sengoi <i>kel'op</i> 'brain'	*q[ɔ]b[u]	*q[a]lu[b](u), AN * <i>qulu[h]</i> 'head'
*ʔ Mundari (P276) <i>tut ka</i> 'back of neck or head', VN <i>ót</i> [ʔɔt] 'nape of neck', Bateg Deq <i>nut</i> 'neck'	*(tu)(n)ʔud	AN *[t]uʔud 'knee'
*b Santali (P206) <i>bɔhɔ</i> , Chrau <i>vôq</i> 'head', Sengoi <i>kel'op</i> 'brain'	*q[ɔ]b[u]	*q[a]lu[b](u), AN * <i>qulu[h]</i> 'head'
*d Mundari (P276) <i>tut ka</i> 'back of neck or head', VN <i>ót</i> 'nape of neck', Bateg Deq <i>nut</i> 'neck'	*(tu)(n)ʔud	AN *[t]uʔud 'knee'
*j Santali (P365) <i>janŋa</i> , PW * <i>joŋ</i> , PM * <i>juŋ</i> 'foot, leg'	*(n)jɔŋ	*() dzɔŋ(dzɔŋ), AN * <i>zerɛŋ</i> 'stand'
*g Juang (P180) <i>goneh</i> , Sora <i>jiñi</i> 'tooth', PM * <i>gnis</i> 'canine tooth'	*g(in)i(s)	AJ *[t,C]agi, AN * <i>gigih</i> 'tooth'
*G Kharia (P305) <i>hepa' d</i> 'near', VN <i>hep</i> 'narrow', Chrau <i>gêp</i> 'close'	*(n)Gep(at)	*(e)(N)Gep, PMY *NGep 'narrow'
*s Kharia (P216) <i>sɔʔlui</i> , Mal <i>nsook</i> , Ruc <i>usúk</i> 'hair'	*(n)suk	* <i>busuk</i> , AN * <i>buhuk</i> 'hair'
*x Kharia (P303) <i>ɛŋgam</i> , Ruc <i>ăsam</i> , PMN * <i>mham</i> 'blood'	*(n)xam	*()ntsa[a]mu[ʔ], PMY * <i>ncyam</i> 'blood'
*h Kharia (P371) <i>jokɔr</i> 'snail', Pacôh <i>clo</i> , Sengoi <i>kalo</i> 'type of snail'	*(n)qo[k]h]	*(N)qo[ho] 'shellfish, snail', AN * <i>kuhul</i> 'snail'
*z VN <i>duoi</i> , Khmer <i>kanduy</i> 'tail', Pacôh <i>xoi</i> 'tail, penis'	*u(n)zay	**[q,ʔ]u(n)zəy 'worm, penis', AN * <i>huDay</i> 'worm'
*ɣ Hill Kharia (PK498) <i>jor</i> 'river', Bahnar <i>jroh</i> 'swift (current)', Khmer <i>jroh</i> 'mountain stream'	*[ja]ɣos	* <i>qaɣus</i> , AN * <i>qaRus</i> 'stream, current'

AA	PAA	AT
*l Santali (PK289) <i>lɔ'j</i> , Ruc <i>luc</i> , Nicobar (Car) <i>ku-lɔich</i> 'penis'	* <i>lac[u]</i>	AN * <i>lasu</i> 'genitals'
*j Mundari (P311) <i>hɔɔ</i> 'a man, a Munda', Khasi <i>kur</i> 'a clan', Old Mon <i>kulo</i> 'kinsman'	* <i>(n)quɔ</i>	* <i>() (N) [Go]lon</i> , AN * <i>hulun</i> 'man (human being)'
*r Kharia (P160) <i>gere</i> , Rengao <i>hora</i> , PM *(? <i>a</i>) <i>daa</i> 'duck'	* <i>(da)ra</i>	AN * <i>da[r]ah</i> 'dove, pigeon'
*R Mundari <i>taran</i> 'shoulder' Ruc <i>səŋ cūbrañ</i> 'side', Brou <i>bran</i> 'ribs'	* <i>baRañ</i> or * <i>baRaŋi</i>	AJ * <i>baRaŋ</i> 'ribs'
*w Kharia (PK535) <i>sou</i> 'husband', Katu <i>sasaau</i> 'sister's husband', PS * <i>bnsaaw</i> 'wife's elder brother'	* <i>saw[a]</i>	* <i>(qa)(n)saw[a]</i> 'woman', AN * <i>[q,ʔ]a[c.s]awaʔ</i> 'spouse'
*y Mundari (P31) <i>tayan</i> , Katu <i>bayua</i> , Sengoi <i>bahya</i> 'crocodile'	* <i>buq[aya]</i>	* <i>[(m)b]uqay(a)</i> , AN * <i>b[uq]ayah</i> 'crocodile'
*m Kharia <i>mamu</i> 'mother's brother', VN <i>mū</i> 'old woman', Mon <i>mū</i> 'parent's younger brother'	* <i>()mu</i>	* <i>[(t)a]mu</i> 'grandparent, grandchild', FO * <i>mumu</i> 'grandfather'
*n Kharia <i>nana</i> 'elder sister', Bahnar <i>nā</i> 'parent's older sibling', Sengoi <i>kenan</i> 'female (animals and birds)'	* <i>(na)na</i>	* <i>[(t)i]na(na)</i> , AN * <i>(t)ina</i> 'mother'
*ñ Kharia <i>guñumdag</i> 'mosquito', Katu <i>jamuq</i> 'gnat'	* <i>g[ñ]a]muk</i>	AN * <i>ñamuk</i> 'mosquito'
*ŋ Santali (P365) <i>jaŋga</i> , PW * <i>joŋ</i> , PM * <i>juŋ</i> 'foot, leg'	* <i>(n)j</i>	* <i>()dzəŋ(dzəŋ)</i> , AN * <i>zeŋzeŋ</i> 'stand'

Consonant development is generally less problematic than the vocalic development, but this does not mean that little diversity is to be seen. Some of the AA proto-phonemes have multiple reflexes, e.g. *q > /k, ʔ, g, h, ŋ, 0/, some evidence remarkable splits, e.g. *z > s but *nz > d, *y > r but *ny > g, while the modern reflexes of others, e.g. *x and *R, have not yet been fully identified. */q, G, x, ʔ, l, R, [N]/ have been completely replaced.

PAA probably possessed a large number of consonant clusters, and some of them evidently coalesced as unitary phonemes under as yet unclear circumstances. This group included nasal/oral clusters, such as *mp, *mb, etc., which have both implosive and simple nasal reflexes. In AT and AN, Benedict and others have set up special symbols to denote the reflexes of certain coalesced clusters, but thus far in AA this convention has not seemed necessary. Some of the AT reflexes are thought to have been retroflexes, which Munda possesses in abundance, but the western AA retroflexes do not generally correlate to proto-clusters and are likely the result of Dravidian and/or Indic influence, as has long been suspected.

3.3.4 *Canonic Structure*

Pinnow (1959:457) proposed (C)V(C) and (C)V(C)(C)V(C) as the PAA canonic syllable and word, respectively. He did not permit consonant clusters in any C slot and felt that any such clusters not formed by combination of PAA syllables had developed secondarily through affixation and canonical reduction.

This study indicates that the PAA canonic structure was somewhat more complex than Pinnow thought. Two types of such structure can be distinguished, one a radical type relevant for lexical and grammatical morphemes taken in isolation, the other a compositional type relevant for composite forms containing both types of those morphemes. The first type consists of a canonic syllable, (C)V(C), and at least two canonic roots, (C)V(C) and (C)V(C)V(C), in which consonant clusters were permitted in the C1 and C2 positions of lexical morphemes, but not finally. The other type is described in the following subsections.

3.4 *The PAA Morphological System*

3.4.1 *The Central Issue*

The morphological typology of the modern AA languages evidences a remarkable divergence. The Munda languages possess a synthetic type of system characterized by nominal and verbal complexes composed of a nuclear stem and various additive morphs (Pinnow 1966:179 gives 12 categories of such compositional elements in the verb complex). EAA has an analytic type in which a tendency to one word: one meaning and parallel ordering of sentence components is prevalent; it is most pronounced in such monosyllabic languages as VN (MK). The central issue in AA comparative morphology is consequently to determine which of these types, if either, is representative of the PAA system.

Pinnow and other Austroasianists have discussed the question of external influences since western and eastern AA can be grouped in distinct Sprachbünde, the former surrounded by the synthetic Dravidian and Indic (Indo–European) languages, the latter adjacent to the analytic languages of Sino–Tibetan and AT. However, Pinnow (1966:186) concluded that it must be assumed that the Munda languages have preserved the older state of morphological affairs, and thus far the results of this study tend to confirm that conclusion, even though the portrayal of PAA morphology essayed below differs greatly from Pinnow's concepts of it.

3.4.2 *The Structure of Lexical Forms*

This study has resulted in a view of PAA (and Proto–Austrie) morphology which differs significantly from all past conclusions about this subject. This view is based on a comparison of lexical forms trans–AA and trans–Austrie which indicates that such forms are composed typically of a less variable mono– or disyllabic nucleus and a more variable syllabic or asyllabic set of segments additive to that nucleus. Since the nuclear element tends to be more stable and affected primarily by phonological change, it is assumed to be the remnant (reflex) of the ancient

lexical stem. And since the additive set tends to vary frequently as to composition and structure, having been evidently often subject to replacement or incorporation into stems, it is assumed to consist of reflexes of ancient grammatical morphemes used to inflect those lexical stems.

These assumptions allow the reconstruction of a lexical morpheme which was composed of a nuclear lexical morph (stem), canonically (C)V(C) or (C)VCV(C), and various grammatical morphs, most of which appear to have been single phonemes. The stem was also the lexical root morpheme, but this root may never have occurred as a free form. The non-root grammatical morphemes were affixes which probably existed only as bound forms. They can be grouped into complexes which are discussed in the following subsection.

To demonstrate this concept, the 'red' etymology cited in Sec. 2.3.3 can be used. Comparison of the group of suspected cognates listed below suggests that their nuclear elements are reflexes of a single lexical root, **ʔa[q]* 'red' (some reflexes indicate **ʔa* or **ʔa[g,G,x]*). Some nuclei could be interpreted as reflexes of morphophonically shifted doublets, i.e. **ʔi[q]*, **ʔə[q]*, **ʔu[q]*. This process of morphological derivation is probably reconstructible at the AA proto-level, but vocalic assimilation is a plausible alternative explanation. Other reflexes suggest reduplication of the root, another PAA morphological process which could occur partially or totally, i.e. **ʔa/ʔa[q]* and **ʔa[q]/ʔa[q]*.

In the following listing, the suspected immediate antecedent is reconstructed and its ostensible modern reflexes are cited thereafter.

- **i/ʔa[q]* (Katu *riiq* in *arââk riiq* 'yellow'),
- **a/ʔa/[k/q]* or **a/ʔa/[n/k/V]* (Pacoh *raq* 'yellow', Rengao *rak* 'yellowish', *dum rak* 'brown', Jeh *rak* 'rusty color'),
- **u/ʔa[q]* (Wa Tung Va *rauh* 'red'),
- **[k/ə]/ʔa[q]* (Katu *arââk riiq* 'yellow'),
- **i/ʔa[q]/i* (Sora *jʔe:* 'red'),
- **b/i/ʔa[q]/(lu)* (Bahnar '*brê* 'red', '*bre*' 'very red', '*bri*', '*brïu* ['briw?]' 'ruddy complexion', Katu *breh brel* 'colored'),
- **b/l/ʔa[q]* (Katu *brôq ahââq* 'to blush'),
- **p/[ə]/ʔa[q]/i* (Stieng *porhê* 'red'),
- **p/u/ʔa[q]* (Chrau *prho prôt* 'very red', Thavung *tôh* 'red', Katu *paprô* 'shy'),
- **[k/ə]/ʔa[q]/i* (Tampuan *kh-ray*, Che' Wong *bərtay* 'red'),
- **g/l/ʔa[q]/i* (Sora *gərcj* 'put to shame'),
- **p/i/ʔa/n/[q]/i* (Chrau *prênh* 'shy'),
- **b/i/ʔa/n/[q]* (Rengao *brêng* 'black', Bahnar *brêng* 'rust'),
- **b/a/ʔa/n/[q]* (Jeh *sūk brang* 'black hair', Katu *brôông braang* 'red'),
- **b/l/ʔa/n/[q]* (Katu *brôông* 'red', Bahnar '*brông* 'describes many things that are all red', Riang *roŋ* 'red'),
- **[q]/l/ʔa/n/[q]* (Katu *arôông* 'red hot'),
- **t/u/ʔa/n/[q]* (Pacoh *turông* 'a shade of red'),
- **j/l/ʔa/n/[q]* (Santali *jərcəŋ* 'deep red, scarlet'),
- **ʔa/n/ʔa/n/[q]/i* (Semai *rəŋəñ* 'red'),
- **ʔa/n/ʔa[q]/p[V]* (Jah Hut *rəŋap* 'red'),
- **i/ʔa/n/[q]/alt* (Mundari *jəŋge'd* 'deep redness, crimson'),
- **b/i/ʔa/n/[q]/alt* (Old Mon bi[rket], Semaq Beri *dəŋat* 'red'),
- **ʔa/n/[q]/i/s* (Bateg Nong *rəgih* 'red').

Some of those reconstructions are admittedly quite hypothetical, and it is probable that some of the cited proto–affixes were added in post–PAA times. The function of most of these affixes also cannot be specified at this time. But if PAA was a synthetic and inflecting language, as much of the evidence seems to indicate, such groups of morphologically related cognates are exactly what one should expect to find in its descendant languages.

3.4.3 *The Structure of Grammatical Forms*

Comparison of the AA/AT lexical data makes the following points clear: 1) The lexical root in PAA was inflected by a large number of affixes. 2) In the modern languages, there is frequent alternation of segmental phonemes and combinations of these phonemes within a syllable preceding the suspected ancient stem, within this stem, and in and around the final segmental position of this stem. 3) Correspondence sets formed by those phonemes and combinations thereof often correlate to affixes and para–affixes or segments thereof found in modern AA or AT or both. *Para–affixes are phonological elements which correspond structurally to affixes but do not function morphologically as such.* Many of them are defunct affixes, but no general conclusions can be arbitrarily drawn about their origin as some have done in the past.²¹ 4) Certain phonological developments and irregularities in phonological correspondence can be explained by attributing them to the influence of such affix–like elements formerly present in the above cited environments.

These observations suggest that the PAA grammatical system had three affix complexes in which positional classes of bound affixal morphemes were successively placed in a fixed order before, within, and after the lexical stem.

The Prefix Complex

Canonic (C/)(V/)(R/) from earlier (C/)(V/)(C/). The C(onsonant) prefix class comprised most of the proto–consonants; its members may have functioned as classifiers, verbal markers, and/or pronominal referents.²² The V(owel) prefix

²¹ Pinnow (1959:11ff.) references the works of Schmidt, Sten Konow, et al. in which almost every word is viewed as decomposable into prefixes and stems. In this view, all presyllables are prefixes, all main syllables stems or roots. Pinnow felt that the phonological and structural relationships in AA are much more complicated than such a simple division would imply and expressed doubts about the general accuracy of that analysis.

²² Pinnow (1959:13f.) comments on the possibility that the AA presyllables are in part remnants of an ancient class prefix system like that found in the Bantu languages of Africa, but notes that a typical feature of the latter, concordance, is absent in AA. My analysis suggests that only the first prefix class is the suspected classifier and that it could occur with any word, perhaps indicating concordance after all. Some Aslian languages have a pronominal referent system whereby a referent, e.g. Sengoi *ki*– ‘he, she, it’, is prefixed to nouns and verbs. This usage parallels to some extent the Khasi articles (see footnote 23). Some connections between the first prefix class and the topic and other AT markers reconstructed by Benedict (1990:125ff.) and perhaps the AN four–focus system (Dahl 1973:118f., Benedict 1990:129ff.) must be suspected, but due to

class comprised most, if not all, of the proto-vowels; they may have served as articles, gender markers, and/or spatial or temporal deictics.²³ The R(esonant) prefix class comprised a homorganic nasal and probably the laterals, */l, ɭ, r/; the nasal is evidently the phonologically conditioned reflex of a larger set of consonantal prefixes (including the laterals). This class probably served to verbalize nominal stems, nominalize verbal stems, and/or form derivatives from both types of stems. Some typical examples are given next.

Set 7. The Prefix Complex.

Gloss	Root	Word Form	AA	AT
bone	*qəŋ	*c/i/qəŋ *c/a/n/qəŋ *c/u/qəŋ	Khasi (Amwi) <i>śiaŋ</i> Lamet <i>cəŋʔaaŋ</i> VN (*c/i/həŋ >) <i>xuŋŋ</i>	*(q/)(n)uʔəŋ
I	*qu	*m/i/n/qu *n/qu/a/(n) *a/qu	Pareng (P402) <i>miŋgu</i> VN <i>qua</i> , Thavung <i>kan</i> Jeh <i>au</i>	**ʔa(ŋ)ku
nine	*qəwa	*c/i/qəwa *c/i/n/qəwa *t/i/n/qəwa/s/i	Khasi (Amwi) <i>hunshia</i> VN <i>chin</i> Ngeq <i>takias</i>	*ts[i]/qəwa
feces	*qi	*a/qi/[s] *i/qi *a/n/qi *t/a/n/qi/t *[p/l]/a/qi	MUK (*ʔeh >) <i>è</i> Kurku (P81) <i>i'j' dung'</i> PW *ʔeŋ 'dung' Mon <i>taket</i> 'anus' Katu <i>tataaiq</i> 'anus'	*[(m)pl]a(N)qi, AN *taqih 'excrement'
nail	*lus	*c/a/(n)lus/(i) *c/u/lus/i	VN (*clawh >) <i>trão</i> Jehai <i>cənrəs</i> Semelai <i>cəruš</i>	AN *[c,s]ulu[h] 'fingernail'

The Infix Complex

Canonic /(V)/(R)C/. This complex mirrors the prefix complex and evidently evolved from it through transposition due to such factors as nasalization of the R–

limited information on AN morphosyntax, it would be premature to draw any inferences about their relationship.

²³ A connection between these vocalic morphemes and the demonstrative prefixes of Kharia (Biligiri 1965:65), *i* 'what', *a* 'which' and *u* 'this', and the Khasi "gender" articles (Rabel 1961: 93ff.), *i* [ʔi:] 'diminutive (gender-and numberless)', *u* [ʔu:] 'masculine singular', *ki* 'plural (all genders)', and *ka* 'feminine/neuter singular', is quite possible. The phonetics of the Khasi particles suggests that they are actually reflexes of *q/(i,u)/ and *k/(i,a)/(or perhaps *n/q/(i,a)/). The Khasi articles are repeated before attributes and predicates, which led Pinnow to speculate that they may be evidence of the missing concordance noted in footnote 22. But Pacoh also has possessive forms with *i*, as in *alép* and *ilép* (possessive form) 'sister-in-law', which might suggest that the vocalic particles also served as case markers.

class prefixes. Hence, the prefix and infix classes were probably composed originally of the same morphophonemes. Typical examples follow.

Set 8. The Infix Complex.

Root	Derivative	Root	Derivative
*qar	*q/(i,a,u)/(n)p/ar	Sengoi <i>har</i> , PW *lʔar 'two'	Khmer <i>b̄r</i> , Juang (P49) <i>ambar</i> , Kharia <i>ubar</i> 'two'
*taʃu	*t/(i,a)/(n)p/aʃu	Alak <i>t̄rau</i> , Kurku <i>туру</i> (P148) 'six'	Kharia <i>tibru</i> , Jeh <i>todrau</i> , Semaq Beri <i>t̄m-p̄r̄l̄</i> 'six'
*qoc	*n/q/o/(n)l/oc/(i)	Santali (PK67) <i>ḡoʃj</i> 'dead', Pearic <i>hoic</i> 'die'	Temoq <i>k̄abos</i> 'die', Khmer <i>khmoc</i> 'ghost'
*cit	*c/(i,a)/(n)p/it	VN <i>ch̄at</i> 'narrow', Pearic <i>cit</i> 'tight'	Kharia (P93) <i>japi d</i> 'close eyes', Santali <i>jipi d̄</i> 'close one eye', Jah Hut <i>fiem</i> 'near'
*qə[ɣ]	*q/(a)p/ə[ɣ]	Khasi <i>heɹ</i> , Khmer <i>hoer</i> 'to fly'	Mundari (PK501) <i>apir</i> , Katu <i>par</i> 'to fly'

Cf. AT *(N)qa(m)bar 'twin, double(d), two', *(q)obots 'ended', *[tsu](m)pit 'narrow', *(q)(m)p̄Rl(l(m)p̄R) 'spread (out), fly', Atayalic *ma/taʃu 'six'.

The Suffix Complex

Canonic (N) ... (C)(V)/(C) from earlier *(C)(V)/(C), with ... representing the stem's final consonant. The identity and functions of the suffixes are less well known due to widespread attrition of such affixes at later stages. Quite likely, the suffixes served to denote morphosemantic distinctions roughly equivalent to the nominal cases and verbal declensions of the Indo–European languages. The "infixid" suffix may have been transposed for euphonic reasons or because the other positions were already full.

Set 9. The Suffix Complex.

Modern AA	PAA	PAT
VN <i>guc</i> 'bend down (head)', Khmer <i>kui</i> 'bend', Chrau <i>cungeq</i> 'head to side'	*(n)kuk, *ku/n/k, *ku/n/k[eq]	*[i]ŋkuk, *[i](ŋ)kuŋ 'bend/ bent', AN *biŋkuk 'bent', *[C,i,T]ikuŋ 'bend'
Mundari (PK537) <i>rowa</i> 'soul, spirit', Pacôh <i>rvai</i> 'soul', Kurku <i>rawan</i> 'spirit', Chrau <i>yang ravang</i> 'the spirits'	*r/(a,u)/wa/(i,n), *r/a/wa/n/[q]	*()m/ñ[a]w[a] 'belly, soul', AN *ñawah 'spirit, soul'

Semai <i>laal</i> 'stick out (tongue)', PW * <i>glak</i> , Khasi <i>jliah</i> 'lick', Santali (P14) <i>alaŋ</i> , Pareng <i>laʔu</i> , Thavung <i>layh</i> , Brou <i>liaih</i> 'tongue'	* <i>la/la</i> , * <i>g/a/la/ak</i> , * <i>ji/la/ s</i> , * <i>a/la/(n)k/ u</i> , *(<i>i</i>) <i>la/ s/i</i>	* <i>[b]li[d]aq</i> 'lick, tongue', AN * <i>dilaq</i> , * <i>[z,Z]elaq</i> 'tongue'
Jeh <i>wa</i> 'ape', PW * <i>hwaʔ</i> , MUK <i>wək</i> 'monkey', Sora <i>kǎró</i> , Sengoi <i>rau</i> 'type of monkey', <i>mawak</i> , Thavung <i>jooç</i> , Pacóh <i>adyoaiq</i> [ʔəʔjoa:yʔ] 'gibbon'	* <i>s/a/wa/(q)</i> , * <i>r/(a,u)/wa</i> , * <i>m/(a,u)/walk</i> , * <i>j/(a,u)/wa/ t/s</i>	AN * <i>mawas</i> 'monkey (species)'
Kharia <i>torog</i> 'cry like an animal', Khmer <i>ro</i> 'low', <i>lū</i> 'howl', <i>ralūn</i> 'howling', VN <i>rōng</i> 'trumpet (elephant)'	* <i>lo/(q)</i> , * <i>lo/n/ q</i>	* <i>lo/ oŋ</i> 'cry (of animal)', AN * <i>lulu</i> 'bark, howl'

Integration of the canonic forms taken by the lexical morphs and grammatical complexes produces the maximal canonic words (or word complexes) shown below.

Canonic Root Canonic Word Form

(C)V(C)	(C)(V/)(R/)(C)/(V/(R/)C)V/(R/)(C)/(C)/(V)/(C)
(C)VCV(C)	(C)(V/)(R/)(C)/(V/(R/)C)VCV/(R/)(C)/(C)/(V)/(C)

Hence, the PAA canonic word was maximally four or five syllables long and perhaps even longer if reduplicated stems were permitted in such maximal forms. The issue of compounding and longer forms produced by it is not addressed here.

Although this inductive analysis of ancient AA morphology may be logically correct, the actual existence of such maximal canonic forms is difficult to demonstrate empirically. The diachronic changes operative in AA have tended to destroy most of the suffix complexes and grind down much of the prefix complex structure, and we are allowed to view only the tattered remnants and speculate on how they once fitted together. Then again, the most common word forms in PAA may have belonged to shorter subsets of the above-described maximal canonic form.

In Munda, both tight and loose binding of additive morphs to complex nuclei occurs; hence, the actual morphological structure of PAA may not have been as rigidly formatted as the above-proposed forms would indicate. Some of the prospective affixes may really have been free syllabic morphemes like **qu* 'I' as in **ta/ta/n/qu* 'my grandfather', whence Kharia *tataŋ* 'grandfather'.

4. The Reconstruction of Proto–Austric

4.1 *More Probatory Problems*

To fulfill the theoretical and methodological requirements of confirming the Austric hypothesis, genetic correspondence must be shown to exist between AA and AT and the Austric proto–language must be reconstructed. In general, the first task is easily accomplished, at least to a certain degree, for regular and recurrent phonological, morphological, and lexical correspondence is rather easily established between AA and AT once the discovery problems are sufficiently overcome. The second task is less easily done, but due to the intrinsic nature of linguistic change, the proto–language we are able to reconstruct can be only a model of the real Proto–Austric language based on the subset of the latter’s features which has been retained encoded in the AA/AT genetic correspondence. Hence, the phonological and morphological “systems” presented below can pretend to be the Austric proto–language, even though it is quite certain that they represent only a certain (and perhaps minor) portion of its total real structure.

Despite the progress made thus far in fulfilling those requirements, the status of the Austric hypothesis remains paradoxical because, as indicated in Sec. 2.2.1, the manner in which AA and AT interrelate historically is still very obscure. There can be little doubt that the demonstrable partial correspondence between them connotes common origin or that they converge to a common ancestor. It is less certain just what that ancestor was and whether Schmidt’s term, Austric (or Macro–Austric as Diffloth proposes to reflect the AT–vice–AN exchange), will remain a viable name for it. As a tool of historical research, the comparative method is very powerful but in the present case not very specific, and one can only wonder just where it has led us.

These points have become increasingly important with the realization that the extent of the AA/AN lexical correspondence surpasses all previous ideas about it. As this data is processed, an increasing number of AA reconstructions are shown to be identical or nearly so to the AN proto–forms. While it is to be expected that formal convergence will appear as the comparison brings us closer to the common ancestor of these languages, the prevalence of this formal identicalness is contradictory in view of the chronology and presumed nature of the AA/AT relationship discussed in Sec. 2.2.1. Hence, it is possible that two ancient strata are coming to light in this comparison, one presumably the “real” AA and characterized by such phonological differences that it cannot possibly be borrowed from any stage of AN or AT, the other reflecting an extensive influx of vocabulary from a very early stage of AN and/or AT. In essence, this is Benedict’s substratum theory recast, with the important difference that both super– and substratum were genetically related.

4.2 The Proto–Austriac Phonological System

4.2.1 Canonic Structure

Benedict (1975:149ff., 1990:19ff.) reconstructs for PAT (C)V(C) and (C)VCV(C) as the canonic syllable and word, respectively, and permits consonant clusters to occupy the C1 and C2 (but not the final) positions. Since he distinguishes between two types of such clusters, one being the nasal–increment type (in which the other can be embedded), the syllabic forms are really (N)(C)(C)V(C) and the word forms (N)(C)(C)V(N)(C)(C)V(C).

These PAT canonic forms and those proposed for PAA in Sec. 3.3.4 differ only in the question of how to handle the nasal increments. This study suggests that nasal–oral clusters developed largely through affixation and while it seems unlikely that Austriac did not have such clusters originally, no clear examples of them have been discovered. Hence, the Proto–Austriac canonic structure is assumed to have been identical to that of PAA, the differences in PAT resulting from affixation and composition.

4.2.2 Vowels

The Austriac vowel system is very uncertain, but it seems probable that the simple set, */i, e, ə, a, u, o/, existed at this level.

4.2.3 Consonants

The PAT consonants (Benedict 1975:155, 1990:50–108) are depicted in the right column of the following data table. Changes to the 1975 version made in 1990 include addition of */tš, tš, dz, š/ and confirmation of */ʎ, ś, x/. In contrast to PAT, the PAA consonant system (center column) is less complex. Hence, it appears at present that Proto–Austriac must be reconstructed with the 24 consonants (plus four uncertain ones) shown in the left column. The exemplary lexical sets supporting most of these proto–phonemes are given in Sec.3.3.3; the data table below presents only those sets in which there is some question about the origin of the AT proto–phoneme.

Set 10. The Austriac Consonants.

Proto–Austriac						PAA						PA						
*p	t	[c]	k	q	?	*p	t	c	k	q	?	*p	t		c	k	q	?
b	d	[j]	g	G		b	d	j	g	G		b	d		j	g	G	
	ts												ts	tš	tš			
	ɬ												ɬ	ɬ				
	s		x	h			s		x	h			s	š	ś	x	h	
	z		Y				z		Y				z		[ž]	Y		
	l		[ʎ]				l		ʎ				l			ʎ		
	r		[R]				r		R				r			R		
w		y				w		y				w		y				
m	n	ɱ	ŋ	/		m	n	ɱ	ŋ	[N]	/	m	n	ɱ	ŋ	N	/	

PA	PAA	PAT	Modern AA	PAA	PAT
*[c]	*c	*c	VN <i>châm</i> , Pacôh <i>dyúm</i> ‘dip’, Sengoi <i>cum</i> ‘set (sun)’	*(n)cəm	*cəmcm ‘dip’, AN *k’əm’əm (D), *cemcem (L) ‘dampen’
*[j]	*j	*j	Riang <i>bus</i> ‘foam’, Thavung <i>buuc</i> ‘make bubbles’, Semai <i>buus</i> ‘bubbles’	*bu[c,j](i)	** (m)pujaq, AN *bucaq ‘foam’
*tʰ	*c	*tʰ	Jeh <i>chol</i> ‘look with light’, Katu <i>dyóól</i> ‘light way’, Bahnar <i>plêr–plar</i> ‘shine’	*(n)caR and *cilaR	AK *[tʰ,tsʰ]ilaR, AN *sinaR ‘light’
*tʰ	*c	*tʰ	Pearic <i>khəcay</i> ‘climb’, Khmer <i>chkay</i> ‘stand straight’, Palaung <i>jok</i> ‘lift’	*(n)caka(i)	AK *(ń)tʰaka(i) ‘rise’, AN *[s]akay ‘ascend’
*dʒ	*j	*dʒ	Rengao <i>yâi</i> , PMN *yoy, PS *ñyñooy ‘lip(s)’	*ñjoy	** gu(ń)dʒuy, AN *gu[s,ts]u ‘mouth’
*s	*s	*s	PW *[lɪm] ‘pus’, VN <i>răm</i> ‘conjunctivitis’, (skin) PM *ksaam ‘epidemic’	*(s)Rəm	AJ *gušam, AN *guhām ‘rash’
*s	*s	*ʂ	VN (*ʒut >) <i>thut</i> ‘draw back’, Pacôh <i>xot</i> ‘pick rice by stripping head’, Bahnar <i>hrot</i> ‘pull lice out’	*srut	*[ʂ]urut ‘pull lengthwise’, AN *hurut ‘stroke’
*z	*z	*z	VN <i>lanh</i> ‘cold’, PW *hleŋ ‘cool’, Khmer <i>sreñ</i> ‘to cool’	*(z)leŋ	*(q)z[en]en, AN *Dirin ‘cold’

As the preceding table indicates, a major phonological difference between AA and AT is the absence of affricates, alveopalatals, and palatal sibilants in the former. In Sec. 3.3.3, it was shown that the Austric affricates have shifted to palatal stops in AA. In the data above, one sees that the palatal stops of PAT correspond to AA palatal stops which evolved out of the Austric affricates; hence, it is questionable that Austric had *c and *j. One also notes that the AT alveopalatal and palatal sibilants correspond to sibilant–liquid clusters in AA, some of which also coalesced but probably in the post–AA era. This correlation leads one to suspect that AT */tʰ, tʰ, dʒ/ are reflexes of affricate–liquid clusters; hence, Austric did not have these proto–phonemes and may not have possessed independent palatal stops.

Velar liquids—a very distinctive trait of PAT according to Benedict—have been reconstructed for PAA on the basis of the correspondence of AA *l* and/or *r* to both AT **l* and **R* (or AN **l* and **R* when no PAT cognate is available). Thus far, little regularity has been noted in this correspondence, and the only criterion on which the reflexes of **l* and **R* can be distinguished from those of **l* and **r* in AA is the fact that the velar liquid reflexes frequently alternate as *l* or *r* whereas the simple liquid reflexes apparently do not. Some of the data suggests that the proto–units represented as velar liquids may actually have been clusters composed of liquids and laryngeals, i.e. */ʎ/, ʀ, hl, hr/ and/or */ʎ/, rʎ, lh, rh/. This situation is still unclear, but this possibility has been indicated by putting the Proto–Austric velar liquids in brackets.

It is quite likely that Proto–Austric possessed a large number of consonant clusters, many of which have not yet come into focus. Benedict (1975:171–8, 1990:71–8) has reconstructed about 20 clusters, all composed of stops plus liquids

1990:71–8) has reconstructed about 20 clusters, all composed of stops plus liquids except for one nasal plus liquid, in addition to the nasal–increment type which comprises both homorganic nasal plus single consonant clusters and homorganic nasal plus the above–described clusters. The AA evidence indicates that sibilant plus liquid clusters also existed and possibly that affricate and laryngeal clusters occurred. The laryngeal type is found throughout AA, although often the reflex of non–laryngeal clusters, and less frequently in mainland AT, but apparently not at all in AN. Certain irregular correspondences, such as AT ***bi(n)tuqan*, Kharia *tErgan*, Katu *satur*, Loven *ptuar* ‘star’, suggest the presence of still other clusters.

4.3 The Proto–Austrian Morphological System

At present, it appears that the Proto–Austrian and PAA morphological systems were virtually identical with three exceptions. The third class of prefixes may not yet have become totally nasalized, infixes may not yet have developed, and the infixed suffix may not yet have appeared.

If this analysis is correct, and this remains to be seen, then the morphological system of PAT as thus far reconstructed by Benedict would have to be regarded as already having undergone massive changes which had reduced the complexity of the earlier system and produced a more rigid word structure in which many of the affixes were already incorporated into the word stem and presumably morphosyntactically defunct. This progressive state of developmental affairs is also incongruous in view of the chronology, etc., discussed in Sec. 4.1 above. The trek continues.

5. Conclusion

The empirical evidence presented in the preceding pages makes an overwhelming case for the conclusion that AA and AT do share partial correspondence indicative of common origin. Borrowing by AA, especially from the substratumized AT group postulated by Benedict, is the less likely explanation for that correspondence’s existence. It cannot be denied that AA has borrowed from AT, perhaps extensively and certainly over a very long period of time, but the demonstrated linkage between them in the areas of phonology, morphology, lexicology, and diachronic development is of such nature that it seems best explained by inheritance. It should, of course, be emphasized that the findings presented here represent only a beginning to the enormous tasks of reconstructing PAA, comparing it formally with PAT, and reconstructing their common ancestor.

In closing, it is now abundantly clear that in his assessment of the AA/AT relationship, Benedict was oceanic in being wide of the mark. Not only do AA and AT share a congruency in overall configuration, but there is also massive evidence of a shared core vocabulary, only a small part of which has been displayed here. But if I have come to bruise Benedict for error in his evaluation of the Austrian hypothesis, I cannot depart without also acknowledging the merit of his contribution in reconstructing AT and the gigantic debt this study owes that accomplishment. Although his views on the AT stock have not been so graciously

corroborate the general accuracy and insight of his comparative analysis. And one fact seems above all certain: without Benedict's pioneering work on AT, it is most probable that this reconstruction would hardly have been thinkable, much less feasible, for a long time to come.

REFERENCES

- Anttila, Raimo. 1972. *An Introduction to Historical and Comparative Linguistics*. New York: Macmillan.
- Banker, John and Elizabeth and Mo. 1979. *Bahnar Dictionary, Plei Bong-Mang Yang Dialect*. Huntington Beach: SIL.
- Barker, Milton and Muriel. 1976. *Mường-Vietnamese-English Dictionary*. Microfiche. Huntington Beach: SIL.
- Benedict, Paul K. 1975. *Austro-Thai: Language and Culture, with a Glossary of Roots*. HRAF Press.
- Benedict, Paul K. 1976. Austro-Thai and Austroasiatic. *AS*, pp.1-36.
- Benedict, Paul K. 1990. *Japanese/Austro-Tai*. Ann Arbor: Karoma Publishers.
- Benjamin, Geoffrey. 1976a. Austroasiatic Subgroupings and Prehistory in the Malay Peninsula. *AS*, pp. 37-128.
- Benjamin, Geoffrey. 1976b. An Outline of Temiar Grammar. *AS*, pp. 129-87.
- Bhattacharya, Sudhibhushan. 1966. Some Munda Etymologies. *SICAL*, pp. 28-40.
- Biligiri, H.S. 1965. *Kharia: Phonology, Grammar and Vocabulary*. Poona: Deccan College, Postgraduate and Research Institute
- Blood, Henry F. 1968. *A Reconstruction of Proto-Mnong*. Grand Forks: SIL-University of North Dakota.
- Blust, Robert A. 1970. Proto-Austronesian Addenda. *Oceanic Linguistics*. 9: 104-62.
- Blust, Robert A. 1980. Austronesian Etymologies. *Oceanic Linguistics*. 19:1-2.
- Blust, Robert A. 1983-4. Austronesian Etymologies—II. *Oceanic Linguistics*. 22-3:29-150.
- Blust, Robert A. 1988. The Austronesian Homeland: A Linguistic Perspective. *Asian Perspectives*. 26.1: 45-67.
- Costello, Nancy A. 1971. *Katu Vocabulary*. Saigon: Bộ Giáo-dục & SIL.
- Dahl, Otto C. 1973. *Proto-Austronesian*, Scandinavian Institute of Asian Studies Monograph Series No. 15. Lund: Studentlitteratur.
- Diffloth, Gérard F. 1977. Mon-Khmer Initial Palatals and "Substratumized" Austro-Thai. *MKS* 6:39-57.
- Diffloth, Gérard F. 1980. The Wa Languages. *LTBA* 5.2.
- Diffloth, Gérard F. 1984. *The Dvaravati-Old Mon Language and Nyah Kur, Monic Language Studies I*. Bangkok: Chulalongkorn University.
- Diffloth, Gérard F. 1989. What Happened to Austric? *MKS* 16-17:1-9.
- Ferlus, Michel. 1974. Lexique souei-français. *ASEMI*. 5.1:141-59.
- Ferlus, Michel. 1979. Lexique thavung-français. *CLAO* 5:71-94.
- Fournier, Alain. 1974. Les Khasi, une population Mon-Khmer de l'Inde. *ASEMI* 5.1: 79-95.
- Gradin, Dwight and Thông. 1979. *Jeh Vocabulary*. Huntington Beach: SIL.
- Gregerson, Kenneth and Marilyn. 1977. *Rengao Vocabulary*. Huntington Beach: SIL.
- Headley, Robert K. Jr. 1978. An English-Pearic Vocabulary. *MKS* 7:61-94.

- Henderson, Eugénie J.A. 1976. Vestiges of Morphology in Modern Standard Khasi. *AS*, pp. 477–522.
- Hoeningwald, Henry M. 1950. The Principal Step in Comparative Grammar. *Lg* 26:357–64.
- Huffman, Franklin E. 1978. *The Centrality of Katuic–Bahnaric to Austroasiatic*. Paper presented to the Second International Conference on Austroasiatic Linguistics, Mysore, India.
- Jacob, Judith M. 1974. *A Concise Cambodian–English Dictionary*. London: Oxford University Press.
- Katičić, Radoslav. 1970. *A Contribution to the General Theory of Comparative Linguistics*. The Hague: Mouton.
- Kuiper, F.B.J. 1948. Munda and Indonesian. *Orientalia Neerlandica, a Volume of Oriental Studies*, pp. 372–401. Leiden: A.W. Sijthoff.
- Lindell, Kristina. 1974. A Vocabulary of the Yuan Dialect of the Kammu Language. *Acta Orientalia*, 36:191–207.
- Luce, G.H. 1965. Danaw, a Dying Austroasiatic Language. *Lingua* 14:98–129.
- Matisoff, James A. 1990. *On Megalocomparison*. Discussion Note. *Lg* 66:106–20.
- Means, Nathalie and Paul B. 1986. *Sengoi–English, English–Sengoi Dictionary*, edited by Gordon P. Means. Toronto: Joint Centre on Modern East Asia, University of Toronto—York University.
- Meillet, Antoine. 1967. *The Comparative Method in Historical Linguistics*. Paris: Champion.
- Miller, Carolyn and John. 1967. *English–Brôu Language Familiarization Manual*. Saigon: SIL.
- Pinnow, Heinz–Jürgen. 1959. *Versuch einer historischen Lautlehre der Kharia–Sprache*. Wiesbaden: Otto Harrassowitz.
- Pinnow, Heinz–Jürgen. 1966. A Comparative Study of the Verb in the Munda Languages. *SICAL*, pp. 96–193.
- Pou, Saveros and Philip N. Jenner. 1975. Proto–Indonesian and Mon–Khmer. *AP* 17.2:112–124.
- Rabel, Lili. 1961. *Khasi, a Language of Assam*. Baton Rouge: Louisiana State University.
- Reid, Lawrence A. 1988. Benedict's Austro–Tai Hypothesis—An Evaluation. *AP* 26.1:19–34.
- Sagart, Laurent. 1990. Chinese and Austronesian Are Genetically Related. Paper presented at the 23rd International Conference on Sino–Tibetan Languages and Linguistics, 5–7 Oct. 1990. Arlington, TX.
- Schmidt, Wilhelm. 1906. Die Mon–Khmer–Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv Für Anthropologie, Braunschweig*, new series, 5:59–109.
- Shafer, Robert. 1952. Etudes sur l'austroasiens. *BLSP*, 48:111–58.
- Shafer, Robert. 1966. *Inroduction to Sino–Tibetan, Part I*. Wiesbaden: Otto Harrassowitz.
- Shorto, H. L. 1962. *A Dictionary of Modern Spoken Mon*. London: Oxford University Press.
- Shorto, H. L. 1971. *A Dictionary of the Mon Inscriptions from the Sixth to the Sixteenth Centuries*. London: Oxford University Press.
- Shorto, H. L. 1976a. In Defense of Austriac. *CAAL* 6:95–104.
- Shorto, H. L. 1976b. The Vocalism of Proto–Mon–Khmer. *AS*, pp. 1041–67.
- Skeat, Walter W. and Charles O. Blagden. 1906. *Pagan Races of the Malay Peninsula, Vol. II*. London: Macmillan. Reprinted 1966, New York: Barnes & Noble.

- Smalley, William A. 1961. *Outline of Khmu' Structure*. New Haven: American Oriental Society.
- Smith, Kenneth. 1967. *Sedang Vocabulary*. Saigon: Bộ Giáo Dục. Reprinted 1977, Huntington Beach: SIL.
- Smith, Kenneth. 1976. North Bahnaric Numeral Systems. *ANS* pp. 61-4.
- Theraphan L. Thongkum. 1984. *Nyah Kur (Chao Bon)-Thai-English Dictionary*, Monic Language Studies 2. Bangkok: Chulalongkorn University.
- Thomas, David D. 1964. A Survey of Austroasiatic and Mon-Khmer Comparative Studies. *MKS*1:149-63.
- Thomas, David and Thổ Sáng Lục. 1966. *Chrau Vocabulary*. Saigon: Bộ Giáo Dục. Reprinted 1977, Huntington Beach: SIL.
- Watson, Richard and Sandra Watson and Cubuat. 1979. *Pacốh Dictionary*. Huntington Beach: SIL.
- Wurm, S.A. and B. Wilson. 1983. *English Finderlist of Reconstructions in Austronesian Languages (Post-Brandstetter)*. Pacific Linguistics Series C-No. 33.

Received: September 1992

2021 Biltmore Dr,
Fayetteville, NC 28304,
USA