On Megalocomparison

James A. Matisoff

Language, Vol. 66, No. 1 (Mar., 1990), 106-120.

Language is currently published by Linguistic Society of America.

DISCUSSION NOTE

On megalocomparison

JAMES A. MATISOFF

*University of California, Berkeley**

INTRODUCTION: THE WAR OF METAPHORS

1. Joseph H. Greenberg's *Language in the Americas* (*LIA*; 1987) has been greeted with dismay by many specialists in Amerindian linguistics[1] (cf. Chafe 1987, Campbell 1988, Adelaar 1989), and defended by the author in a reply to Campbell in *Language* 65.1 (1989).

There is no denying that G's central thesis—that all the native language families of the Western hemisphere (except for Eskimo-Aleut and Na-Dene) are genetically related, descending from a common ancestor, 'Proto-Amerind', spoken about 11,000 years ago (335)—has a certain romantic sweep to it. Such a far-reaching claim would seem to require pretty convincing evidence to back it up. Yet G expresses his disdain for the conventional sort of historical comparison, so 'obsessionally' preoccupied with sound correspondences and asterisked reconstructions.[2] Instead he values the kind of evidence provided by the more 'powerful' method of 'mass comparison', whereby wordlists and grammatical paradigms from many languages are grokked simultaneously to see whether any root-words or affixes look alike. Languages which share a number of resemblant vocabulary items are then deemed to be genetically related, unless these forms are obvious loans (22–23).

So powerful is this method, G believes, that it yields valid results even with the worst data, and to any desired time-depth: 'The method of multilateral comparison is so powerful that it will give reliable results even with the poorest of materials. Incorrect material should have merely a randomizing effect' (29). Errors do not greatly concern G: 'Although I have exercised great care, it would be miraculous if, in handling such a vast amount of material, there were no errors of fact or interpretation' (1989:112). To this G's critic Adelaar (a Quechua specialist) rejoins that 'This is highly questionable if one looks at the quality of the data G presents where the number of erroneous forms probably exceeds

---

[1] My own ignorance of the Amerindian data (which I have managed to retain despite my Berkeley formation) ensures a certain objectivity, which is perhaps why I have been asked to contribute this Discussion Note. At least I cannot be accused of an axe to grind based on an 'accident of expertise' (*LIA*, p. 4)!

[2] Until very recently linguists in the People's Republic of China have held a similarly unflattering view of the reconstructive enterprise, likening it to 'painting ghosts' (*huà guǐ*)—i.e. trying to lend a specious reality to something imaginary.

that of the correct forms' (1989:253). The consensus among professional Amer-
indianists seems to be that G has NOT in fact 'exercised great care' in the
selection or utilization of his materials, and that he has ignored the results of
the best recent research on many topics.

The most damaging accusation against *LIA* is that its methodology and data
are so inadequate that they are incapable of distinguishing between similarities
that might be due to genetic relationship and those which are due to pure chance
(see §2.4 below).

Yet in a way all this criticism of *LIA* misses the point. Some scholars are
impervious to criticism because they are totally convinced that their underlying
thesis is correct, regardless of any disconfirmatory factual 'details' that might
be wrong.[3] In G's case this conviction derives from a larger belief system that
far transcends the narrow confines of the New World, but embraces our whole
planet and our whole species. Although G provisionally recognizes about 15
'major linguistic families' in the world (Table 11, 337), it is clear that in his
heart he believes he can show that all human languages descend ultimately
from a single common ancestor (62), since 'there is no theoretical limit to the
depth at which classification can be carried out when the number of languages
examined is large' (28–29). *LIA* ends on this very note: 'The ultimate goal is
a comprehensive classification of what is very likely a single language family.
The implications of such a classification for the origin and history of our species
would, of course, be very great' (337).

Proto-Amerind is only the tip of the iceberg. We are promised another book
shortly (G, forthcoming) that will prove that Indo-European itself belongs to
a much vaster language family called 'Eurasiatic', which includes Japanese,
Ural-Altaic, and Eskimo-Aleut (332). For many linguists, such views fall more
into the category of religious beliefs than scientifically testable hypotheses,
about on a par with claims that 'all languages have the same underlying deep
structure' and 'the position of the stars at the moment of our birth determines
our character'.

Since *de dogmatibus non disputandum est*, we must concede that *LIA* cannot
really be judged by conventional criteria. Much of the controversy surround-
ing the book is better considered to be a dialectic between sets of opposing
personality types and intellectual bents: battles of images, metaphors, tem-
peraments.

Certain traditional dichotomies ('theory-oriented' vs. 'data-oriented' or 'lum-
pers' vs. 'splitters') do not quite succeed in capturing the differences between
the opposing camps. G has his theories as well as his data; and many of G's
critics are themselves lumpishly inclined, and would like nothing better than
to succeed in demonstrating a genetic relationship between two families pre-
viously thought to be unrelated. I have made a slightly more nuanced tripartite
distinction among three sorts of linguistic comparison ('micro-', 'macro', and
'megalo-'), according to the putative closeness or remoteness of the genetic

---

[3] In Oriental linguistics, one thinks of such great figures as Bernhard Karlgren and Paul K.
Benedict.

relationship being investigated, and have observed that 'different species of maniacs' are attracted to each variety (Matisoff 1976:258; 1979:38).[4] G seems to scorn those who work at sub-megalo levels of classification as focusing 'on a limited group determined by accidents of expertise' (4).[5] He has no taste for tiny details.

Still another relevant dichotomy is between armchair- or library-scholars on the one hand (let us call them 'fauteuillistes' for short) and fieldworkers on the other. G is definitely a library-type scholar, justifiably proud of his Sitzfleisch in having gone through 'well over 2000 sources' during the preparation of *LIA*. The fieldworker is only too bitterly aware of the inadequacies and messiness of his own material, how the accidents of elicitation cause one form rather than another to be recorded in one's notes during a too-brief stay in some village. The armchair linguist sees only the list of words on the page, authoritative and invariant, artificially neat.

All this having been said, if I were to pick a single word to describe G's apparent motivation in doing megalocomparison, it would have to be COLUMBI-CUBICULOMANIA—a compulsion to stick things into pigeonholes,[6] to leave nothing unclassified. G gives the impression that the highest intellectual activity is the act of classification itself, regardless of the nature of the evidence upon which the classification rests: 'Basically, the wrong question has been asked, namely, when are languages genetically related? ... What should be asked is, how are languages to be classified genetically?' (3). He is fascinated with the astronomically high number of mathematically possible ways there are to classify a relatively small set of objects (6). Once a single overarching classification has imposed order on this chaos, the really interesting part of the work is apparently over. Let others worry about the trivial details of classification at lower taxonomic levels.

METHODOLOGICAL RIGOR IN A MESSY WORLD:
SHOULD MEGALOCOMPARISON BE EASY OR HARD?

2.1. MIXED LANGUAGES, DIFFUSION, COMPLEXITY, AND MESSINESS. G refers with scant respect to believers in mixed languages like C. Loukotka (38), and seems uncomfortable with notions like linguistic areas, or the diffusion of lexical and grammatical traits across genetic boundaries. For anybody who has

---

[4] 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 relationship, where sound-correspondences are not regular and putative cognates are few, so that chance rivals genetic relationship as the explanation for perceived similarities.

[5] The micromaniac might retort that real expertise in a relatively limited area is the best training for more farflung endeavors, and that G would have done well to acquire such expertise before writing a book like *LIA*.

[6] I plead guilty to this awful neologism (< Lat. *columba* 'dove; pigeon' and *cubiculum* 'small compartment'), though perhaps 'pre-columbicubiculomania' is more appropriate for the prehistoric New World. Personally I have come down on the side of those who prefer the continuum to the pigeonhole: 'Pigeonholing is merely a heuristic attempt to make a continuum appear discrete' (Matisoff 1976:258).

worked in a hothouse diffusional environment like Southeast Asia, it seems self-evident that big chunks of even a language's basic vocabulary[7] and grammar can quickly and easily be remodelled under the influence of any languages with which it comes in contact, whether or not such contact languages were genetically related to it in the first place.[8] Borrowing, conflation/contamination/blending, folk etymology, semantic slippage, calquing, backloans—all kinds of messy phenomena complicate matters. These are not marginal, nor, I submit, anything to be ashamed of.[9]

It is not even necessary to travel to exotic climes to convince oneself of the rapidity with which the genetic pathways of words can become obscure. It can happen even with neologisms, before one's very eyes, in less than a generation, in one's native language. How many opinions are already current among the American Volk on the etymology of, say, *T-shirt*?[10] If we are so unsure about T-shirts, how can we be so sure about etymological connections that purportedly go back 11,000 years?

2.2. STRICTNESS VS. LAXITY AND THE CONSEQUENCES OF ERROR. G likes to caricature the traditional historical linguist as a timid soul who dares not venture beyond the lowest level of comparison, considering languages 'pairwise', only two at a time, afraid to tackle huge piles of languages at once. In fact, however, sober-minded scholars have shrunk from megalocomparisons not because they are so difficult, but because they are so easy.[11] When the number of languages

---

[7] An extreme case of relexification is the Bai (= Minjia) language of NW Yunnan, spoken by well over a million highly Sinicized people. While Bai definitely belongs somewhere in the Sino-Tibetan family, its precise genetic classification is difficult, since some dialects have undergone replacement of as much as 75% of their lexicon by Chinese (see Zhao, forthcoming). Sorting out the inherited vs. the borrowed components of the Bai lexicon is an intricate business. Needless to say, languages do not have to be genetically related for one to relexify the other, if the contact is intense enough.

[8] It is as obvious to the areal linguist as to the creolist that no aspect of phonology, grammar, or semantics is immune to borrowing or contamination. Asian languages furnish many spectacular examples, e.g. the monosyllabization and tonalization of the Austroasiatic language Vietnamese under Chinese influence; the acquisition of SVO order by the Karenic Branch of Tibeto-Burman (under Tai and Mon influence); the loss of tone in frontier Chinese dialects of Gansu under Altaic influence; the development of elaborate verbal morphology in many minority languages of India under Indo-Aryan influence; and parallel grammaticalizations of root-verbs to aspect-markers throughout the region (Matisoff 1988).

[9] In a well-written but rather sarcastic passage (11–16), G lists ten 'options' comparativists use to salvage an etymology which shows an irregular sound correspondence, of which one is invoking 'dialect mixture' (16). By rubbing the noses of comparativists in irregular cases (as if there were something shameful or unnatural about them), G is casting aspersions on the comparative method itself, thereby justifying his disinterest in regular correspondences or asterisked reconstructions. Missing from G's discussion is the obvious point that only the idealized assumption of regularity makes it possible to identify 'irregularities' in the first place.

[10] The position of the sleeves with respect to the torso resembles the letter T, and that is apparently the 'true' etymology. Yet the variant spelling *tee-shirt* suggests a sporting connection with golf tees (my own previous belief). Still other members of the Volk (like my wife) believe the T stands for *tennis*.

[11] Campbell uses similes like 'shooting fish in a barrel' (603) or 'playing poker with all the cards wild' (605), and would probably not object to 'taking candy from a baby'.

being considered is large, when their relationship (if any) is remote, and the criteria for sound correspondences are lax, it is not very hard to find 'phono-semantic lookalikes'—forms which more or less resemble each other both in sound and in meaning.

G's faith that multiple errors will somehow have a randomizing effect makes it easy to ignore criticism on any particular point. Other metaphors may be invoked, however—instead of 'cancelling each other out', errors may be COM-POUNDED. The controllers of Voyager II didn't dare trust that any little errors in its trajectory would 'cancel each other out'. A tiny error on the cosmic scale grows to parsecs before you know it. Other images for the dynamics of multiple errors include the Two Drunks Supporting Each Other, the House of Cards, and the information-processing slogan 'Garbage in, garbage out'. (We shall return to this issue in connection with 'teleoreconstruction' in §4.2.)

2.3. SUBGROUPING WITHOUT RECONSTRUCTIONS. Many still remember the brilliant plan proposed by Sen. Aiken (R-Vt.) to bring the Vietnam War to an end. 'Let us,' said the Senator, 'simply declare that we've won, and bring the troops back home.' In a way G's thesis, if accepted, would have a similar effect, by cutting off any further discussion about WHETHER any Amerindian language was related to any other. The only interesting issues remaining would be HOW a language fit into the family as a whole. SUBgrouping then becomes the only meaningful activity for Amerindianists.

But what has been gained? G admits cheerfully enough that 'the internal subgrouping of the 11 Amerind groups remains largely unknown' (64). Isn't having no idea of HOW the language families of the Western Hemisphere relate to each other pretty similar to not knowing WHETHER they are related to each other or not? And how will Proto-Amerind ever be subgrouped without reconstructions of the individual subgroups?

It has often been pointed out[12] that genetic nonrelationship can never be proven. G calls this profound observation 'uninteresting', and counters by saying that you CAN disprove erroneous claims of relative degrees of relationship, e.g., 'Nahuatl is closer to Swahili than to Pima', since 'Nahuatl and Pima ... belong to a valid genetic group (... Uto-Aztecan) that does not include Swahili' (5). Yet one can easily imagine a more subtle Gedankenexperiment, where the nature of the genetic grouping, if any, is not obvious in advance. Suppose we take three languages (A,B,C), all quite different from each other (though similarities exist between any dyad of them, as well as among all three), spoken in widely separated geographic areas, and none of whose further genetic affiliations are known with certainty. How can we tell whether A is more like B than A is like C or than B is like C? G seems to imply that mere eyeballing is always enough to determine the subgrouping of entire languages, just as it suffices to establish the relative phonological resemblance among given lexical items: 'Is a form A more like B than it is like C? Given, for example, *pan / fan / ezuk,* who would hesitate?' (5). But how can this bravado be sustained

---

[12] For instance by the great French comparativist of SE Asia, André-Georges Haudricourt.

in the case of even slightly more difficult examples, such as *pan* / *fan* / *vin* or *bol* / *biŋ* / *bo*? Or *èr* / *erku* / *duo*?[13] Doesn't one sometimes have to hesitate a bit?

G, perhaps making a virtue of necessity, belittles the importance of phonological reconstruction: '... anything approaching a complete and highly convincing reconstruction on the basis of recurrent correspondences is in general possible only with languages so closely related that it is unnecessary anyway' (33). What then is 'necessary' and what is sufficient in historical linguistics? For G, creating the overarching classification is the only intellectual act of any interest. The details can be left to the drones working amid the pigeon-droppings in the lower cubicles.

Quite a different mind-set is displayed in the *Dravidian Etymological Dictionary* (*DED*) of Burrow & Emeneau (1961). Although this great work contains almost 6000 cognate sets, it does not offer any reconstructions either—but this is simply because, after many decades of toil, the authors felt they had not yet resolved all the problems in the reconstruction of the vowels.[14] In every other respect the *DED* presents a maximal contrast to *LIA*: the sets of forms presented are truly cognate, and any irregularities are at least identified and possible explanations are suggested. Errors are at a minimum. Of course the time depth of Proto-Dravidian is much shallower than G's Proto-Amerind, and the validity of the Dravidian classification could just as well have been established by looking at 100 sets instead of 6000, so from G's point of view the whole book was unnecessary in the first place—though it is sure to come in handy when G tries in some future book to subsume Dravidian under a larger family!

Given G's aversion to reconstruction, it is curious that he still operates with the notion of COGNATES: 'The broad approach advocated here does not require the reckless positing of risky and uncertain etymologies. All that is needed is to show decisively more cognates than those of any rival hypothesis' (37). But how are we to decide which etymologies are totally convincing and which are 'risky and uncertain', if we arrived at them by naked eyeballing in the first place? How is a cognate different from a chance lookalike? How do we judge among competing etymologies for the same word? Do lots of erroneous cognate identifications outweigh a few solid ones? If we put in ENOUGH garbage does good stuff eventually come out?

2.4. DISTINGUISHING CHANCE SIMILARITIES FROM RELATIONSHIP. G seems to be tilting at a straw man when he attacks '... the simplistic, but widely held, assumption that after a not very long period the resemblances between two related languages become indistinguishable from chance. This would be true only if there were just two languages in the world' (1989:109). We are in fact supposed to be talking about VERY long periods indeed! Nobody is claiming that all traces of genetic relationship can be obliterated in a short time.

Everybody has a favorite list of accidental lexical resemblances (e.g. Thai

---

[13] These are the Mandarin, Armenian, and Latin forms for 'two'.

[14] Personal communication, M. B. Emeneau (September 1989).

faj, taaj, rim / Eng. *fire, die, rim*). Manipulators of these curiosities may now add G's 'final piece of preliminary evidence' (57) for the reality of the Amerind family, the 'Amerind etymology "hand, give, take"' (Table 9, 58), containing beautiful proto-Sapiens forms like Mayna (Andean) *mani* 'arm', Akwa'ala (Hokan) *man* 'hand', etc. Many scholars have attempted reductiones ad absurdum of megalocomparison by taking any two languages at random and finding large numbers of 'cognates' between them, often to very amusing effect (e.g. Callaghan & Miller 1962). Campbell (1988:602–3) tries the same gambit with Finnish and Amerind, but G is distinctly not impressed: 'I would never compare Finnish in isolation. If Finno-Ugric and the larger Uralic group to which it belongs were not already recognized, I would have discovered them' (1989:111). Can it be that G is missing the point here? THIS CHANCE COMPARABILITY OPERATES EQUALLY WELL BETWEEN ANY TWO PROTO-LANGUAGES CHOSEN AT RANDOM. One could undertake systematic comparisons between, e.g., Proto-Sino-Tibetan and 'Proto-Amerind', and find a dozen or so really snazzy look-alikes. (I will refrain from listing the ones I have found, for fear of providing grist for anybody's mill.[15])

G does seem to recognize the difference between accidental and genetic relationship: 'Of course there are occasional resemblances [between Amerind and Na-Dene and Eskimo-Aleut], attributable either to accident or to common membership in some still deeper grouping' (61). But he then goes on to give a wildly fantastical example (he calls it 'instructive') of an item which seems to him to be a good candidate for the Proto-Sapiens lexicon: an etymon meaning 'hand; finger; to point (with the finger); one', attested by Amerindian forms like Karok *ti:k* 'hand; finger', Yagua *tiki* 'one', Eskimo *tik-iq* 'index finger', as well as by Proto-Indo-European **deik-*'point', Proto-Sino-Tibetan **tik* 'one', and Nilo-Saharan forms like Maba *tek*, Fur *dik* 'one'. 'It may be more widely distributed for aught I know ... it is likely that some of these [resemblances] reflect a common inheritance from a very extensive family—which may even be proto-Sapiens' (62).

To soften up the reader for the lexical presentations to come, G first devotes some space to morphological comparisons (44–57), discussing 'a few widespread grammatical markers ... often involving shared irregularities' (44), e.g. the pervasive Amerindian pattern of pronominal nasal morphemes, *-n-* for 1st person and *-m-* for 2nd person. He returns to this topic in his Reply to Campbell: 'This distribution cannot be explained either by borrowing or chance. The borrowing of first- and second-person pronouns is very rare' (113). Again one wonders how G can be so sure about rejecting alternative possibilities. As far as chance goes, one could point to the quite similar and equally pervasive pattern of nasal morphemes found in those Tibeto-Burman languages with pronominal agreement systems, i.e. *-ŋ-* for 1st person, *-n-* for 2nd person—reduced

---

[15] List furnished upon request to adults 18 years old or over. The amateur linguist and pioneer Tibeto-Burmanist Robert Shafer (1952) seriously entertained the possibility of a relationship between ST and American Indian languages, and Edward Sapir carried on a lively correspondence on this subject with Berthold Laufer in 1920.

from the full forms of the personal pronouns, PTB *ŋa-y ('I; me') and *naŋ ('thou; thee'), respectively. (The correspondences Amer. -n-/TB -ŋ- and Amer. -m-/TB -n- even show a nice parallelism—in both cases the Amerindian nasal is 'one place further to the front' in point of articulation than the ST one!) As for the 'unborrowability' of 1st or 2nd person pronouns, one need only point to modern Thai students' slang, where the English-derived pronouns ʔaj 'I' and juu 'you' have steadily been gaining ground, partly out of exotic chic and partly as a welcome way of avoidng the elaborate distinctions built into the native pronominal system (see Cooke 1968:11, 15). At a more ancient time-depth we could mention the polite 1st-person pronouns for male speakers in Burmese (cun-to; 1st syll. < Old Bs. kywan) and Cambodian (khñom), which both originally meant 'slave'. Here the cultural notions of hierarchy that diffused throughout 'Hinduized' SE Asia are responsible for the parallel grammaticalizations of 'slave' to '1st person pronoun'—though the Burmese and Khmer words themselves are of course not cognate, and one is not borrowed from the other.

In fact, the more languages one looks at, the more accidental resemblances one will find in the phonological shapes of semantically similar functors and affixes—perhaps even more accidental similarities than in the case of root morphemes. It seems impressionistically that formatives with long-lived sounds like nasals and -s- are particularly abundant in the world's languages—perhaps because these sounds are best equipped to withstand the attrition brought about by high textual frequency.[16]

In any event, functors or grammaticalized morphemes are no more immune to accidental phonological similarities than are root morphemes.

2.5. DISTINGUISHING AREAL FROM GENETIC RELATIONSHIP. G's methods lack the subtlety to distinguish similarities that are due to typological and areal factors from those that reflect genetic relationship. One can only imagine what he would do when confronted with a complex linguistic area like Southeast Asia, a region that is home to several quite distinct but highly ramified language families that have undergone mutual influence for millennia. One of the most striking areal features of 'Sinospheric' SE Asian languages is monosyllabicity and elaborate tone systems.[17] Tai, Hmong-Mien (Miao-Yao), and Vietnamese all have Chinese-type tone systems and thoroughgoing monosyllabicity, and all share a good-sized lexical component (including some core vocabulary) which corresponds more or less regularly in consonants, vowels, and tones. If this were brought vividly to G's attention he would probably accept it at once as conclusive evidence for the genetic relationship of all these languages (especially since he would have little to fear from contradictory morphological evidence, in view of the rudimentary morphology of Sinospheric languages).

[16] I. Goddard invokes a 'universal tendency for primary grammatical morphemes to consist of a single, unmarked (phonetically commonplace) segment' to explain the widespread appearance of -n- as a 1st person marker (1986:202).

[17] I refer to the Chinese and Indian areas of linguistic / cultural influence in Southeast Asia as the 'Sinosphere' and the 'Indosphere'.

Yet there is overwhelming evidence that Vietnamese belongs to the Mon-Khmer (Austroasiatic) family (quite unrelated to Chinese), while Tai and Hmong-Mien, while perhaps ultimately related to each other, are more plausibly grouped with Austronesian than with Chinese.[18] Austronesian and Austroasiatic have a number of apparently deep grammatical features in common, including a causative morpheme in -*p*- and infixation as an ancient morphological process, yet share practically no core vocabulary. In any case, the chief interest of SE Asian diachronic linguistics lies in attempting to unravel the threads of diffusional vs. genetic vs. accidental factors in the rich areal fabric. Were we to cut the Greenbergian knot and assume from the outset that all these language families are ultimately related, we would be led into terrible dead ends, and much of the interest would go out of this fascinating field.

SOME PROPOSALS FOR REMOTE LINGUISTIC RELATIONSHIPS IN ASIA:
THE CASE OF JAPANESE

3. For G, 'GENETIC RELATIONSHIP IS PLAINLY TRANSITIVE, so that if Baltic is related to Slavic, and also to Germanic ... then Slavic must be related to Germanic ...' (26; emphasis mine). Sometimes, however, this principle seems to amount to accepting and combining all genetic claims that have ever been made by anybody: if Linguist A suggests that language families X and Y are related, and Linguist B suggests that families Y and Z are related, that means ipso facto that X and Z are related—regardless of the independent validity of either dyadic grouping.[19] It is as if any two genetic hypotheses involving the same language automatically reinforce each other, permitting G to make ever broader groupings.

Yet as the proverb goes, a chain is no stronger than its weakest link. If either the XY or the YZ grouping is shaky to begin with, the XYZ chain can be no stronger. Consider, e.g., Japanese—arguably the world's most culturally important language whose genetic affiliations are still controversial. At least 7 theories have had their adherents: (1) HELIOPARTHENOGENESIS. The Japanese people descend from the Sun Goddess and were dropped down onto the Japanese islands along with their language at the beginning of time. (2) JAPANESE-DRAVIDIAN. The respected Japanese linguist Ohno Susumu has written a book (1980) attempting to demonstrate genetic relationship between Japanese and Tamil.[20] (3) JAPANESE-TIBETO-BURMAN. The eminent Japanese linguist Nishida Tatsuo has argued rather for a genetic relationship between Japanese and another SOV language family, Tibeto-Burman (1978), even attempting morpheme

---

[18] See Benedict 1975, Hashimoto 1976; for opinions to the contrary see Wulff 1934, Li 1976, Yan 1983, Dong et al. 1984.

[19] See, for instance, *LIA* p. 74: 'As is so often the case, they were both right ... If we combine the two chains of suggested relationships, we have, then, Panoan, Tacanan, and Moseten on the one hand and Mataco, Guaicuru, Charruan, Lule, and Vilela on the other.'

[20] So obvious is this relationship to Ohno that he does not feel the need of using Proto-Dravidian reconstructions, but rather compares Japanese to Tamil directly. It must be admitted that the typological similarities between Japanese and Dravidian are indeed striking, and there are even a fair number of phonological lookalikes in grammatical morphemes.

by morpheme comparisons of TB post-verbal particle strings with Japanese inflectional suffixes. (4) JAPANESE-AINU-KOREAN-CHINESE. Johannes Rahder spent years compiling pages of putative Chinese, Korean, and Ainu cognates to Japanese words (see, e.g., 1956/1959), comparing forms from modern Chinese dialects with selected syllables of semantically similar Japanese words. (5) JAPANESE-KOREAN-ALTAIC. Building on the work of S. Martin, who compared Japanese and Korean (1966), R. A. Miller develops this genetic theory in a supremely confident book with the question-begging title *Japanese and the Other Altaic Languages* (1971). (The membership of Korean itself in the Altaic family is still highly controversial.) (6) JAPANESE-AUSTRONESIAN. The simple CVCV structure of Japanese, as well as the geographic unity of the Japan-Ryukyuan-Taiwan string of volcanic islands, has long suggested a relationship between Japanese and Austronesian (Labberton 1924, Solomon 1974, Kawamoto 1977/1978), and this idea has recently been elaborated in stupefying detail by Benedict (forthcoming), who now considers Japanese to belong to his 'Austro-Tai' family, which also includes Tai and Hmong-Mien (see below). (7) G believes that Japanese is but one of 9 subgroups of EURASIATIC, a great family which also includes Eskimo-Aleut and Indo-European.

Leaving aside (1) for the moment, we may apply G's 'principle of transitivity' to the above theories, removing any apparent contradictions among them by assuming they are ALL valid—i.e. that Dravidian, Sino-Tibetan/Tibeto-Burman, Altaic, Ainu, Austro-Tai, and Indo-European are all related to Japanese and to one another. We can be sure that it is by such leaps of transitivity that G will progress toward his ultimate goal of linking up all human languages into the Sapiens family.

GREENBERG AND BENEDICT COMPARED AND CONTRASTED

4. It is instructive to compare G to the foremost megalocomparativist now working on East and SE Asian languages, Paul K. Benedict (B), whom G cites with approval (62, 336). B, like G, views himself as something of an outsider who uses common sense to beat the specialists at their own game. Both are armchair linguists and voracious wordlist consumers.[21] Both have anthropological backgrounds and were decisively influenced by A. L. Kroeber. And both are always absolutely convinced that they are right. Yet while G's chief intellectual pleasure seems to lie in the act of classifying or pigeonholing, B's central metaphor is rather the jigsaw puzzle. His favorite cry of etymological triumph is 'Fits perfectly!' (Never mind if a piece of the puzzle has to be rammed in by force now and then.)

The crucial difference between G and B is that B is in fact a brilliant practitioner of the traditional comparative method. When the scale of comparison is not too vast—i.e. the macrolevel as represented by Proto-TB or Proto-ST— B's instincts are sure, and he is usually right. The problems arise at the 'megalolevel', as represented by Austro-Tai-Japanese (Benedict 1975; forthcoming). Here B's desire to make all the puzzle pieces fit together neatly causes

---

[21] B is said to be able to memorize a wordlist after a single scanning.

him to attempt the impossible: to give the impression of micro-rigor when dealing with (even poorly recorded) megalo-data. Even at the megalolevel B pays homage to the traditional trappings of the comparative method. He claims he is establishing 'regular' correspondences, he provides asterisked reconstructions bristling with parentheses, slashes, and brackets. Sometimes these are so complex and arbitrary that one feels like calling them 'pseudo-micritizing devices'—notational attempts to make the speculative seem rigorous. To put it another way: G sweeps difficulties under the rug, while B offers an instant 'explanation' for any difficulty. B can only cry 'Touché!' to G's remark, 'For those who see reconstruction as proof, there are so many quite legitimate ways of explaining what are apparently irregular correspondences that there is no empirical way of disproving a reconstruction' (10).

4.1. PSEUDO-MICRITIZING DEVICES. One way of ensuring apparent 'regularity of correspondence' is to reconstruct proto-forms that are so complex canonically (e.g. containing long consonant clusters, or even several syllables) that no given combination of proto-entities is likely to recur very often—thus obviating counterexamples. I like to call this 'proto-form stuffing'.[22]

In B's conception, the atonal polysyllables of Proto-Austro-Tai (PAT) were simplified in the Tai-Kadai and Hmong-Mien branches into tonal monosyllables, presumably under the overwhelming areal influence of Chinese. As one can imagine, with these complex polysyllables for proto-forms and these simple monosyllables for reflexes, the etymological possibilities are endless.[23]

If we add to this a rich apparatus of bracketings to optionalize virtually every portion of the proto-form, we are then home free. Thus a huge panoply of reflexes can be accommodated under B's PAT etymology RED/DARK-COLORED/ REDDEN/SHAME(D), reconstructed thus (1975:361): *iɣaŋ; *(q/)bliɣaŋ; *(q/)m/ iɣaŋ; *iɣa(ʔ); *m(a)/iɣa(ʔ); *[i]ɣa(ʔ)/i; *m(a)/iɣa(ʔ)/i. There is, for instance, no problem in identifying as cognate Proto-Mien *ʔñay 'ashamed' and Proto-Tai *hmliaŋ 'rust', the former via *q/mryay and the latter via *q/mriyaŋ < *q/ m/(i)yaŋ.[24] Japanese, with its short and phonologically simple morphemes, is also gratifyingly easy for B to fit into the Austro-Tai picture, as with HAIR/BAST/ HEMP/BEARD/EYEBROW (B forthcoming: 233–5): PJse/AT *[qa-](n)tsa(m)bo[t,c] 'bast; hair; feather > pre-Jse. *[q]a-sa[wo] > Jse. asa 'hemp'; HAND/FIVE (235– 7): PJse/AT *(ka-)lima > pre-Jse. *yi[ma] > Jse. i (as in i-tu-tu, 'with reduplicated numeral suffix').

The height of Benedictine megalocomparative ingenuity is reached in the

---

[22] Weidert 1987 furnishes many good examples of this strategy within the confines of the monosyllable, e.g. in his PTB reconstruction *mrgsla for 'moon', where the monstrous cluster mrgsl- is set up to account for a unique correspondence of initial consonants—despite the fact that this flies in the face of the proto-syllable canon, and in general shows a lack of 'proto-Sprachgefühl' (see Matisoff 1982). (The standard reconstruction of this etymon is *s-la or *s-gla.)

This is not to deny that there may be cases of 'unique but regular' correspondences, as G insists (9). Thus one might well accept B's revision of the reconstruction of the initial cluster of PTB 'blood' from *s-hw- to *s-hyw-, since this does not strain the proto-syllable canon too much, and a parallel cluster *kyw- is attested in the root for 'yam' (1972:51).

[23] Cf. Campbell's discussion of 'short forms and unmatched segments' (1988:600).

concept of SPLIT COGNATES, i.e. cognates that have reflexes of at most one given proto-phoneme in common, since they descend from different syllables of a polysyllabic etymon. This is a powerful reconstructive tool indeed[24], used to good advantage in etymologies like PAT *[wa]kḷəwm[a] 'dog' > Proto-Tai *hma, but > Proto-Hmong-Mien klu (B 1975:272–3); or PJse/AT *(m)ba(ŋ)ʔiwak, which yields the Japanese 'split doublets' uo 'fish' (< *uwo < *iwo < *ʔiwak) and wani 'crocodile' (<*baŋi[wak]) (219–20).

4.2. 'TELEORECONSTRUCTION' AND CONFLICTING MEGALOGROUPINGS. G claims that 'The validity of Amerind as a whole is more secure than that of any of its stocks' (59)[25], and maintains that phenomena which are inexplicable on a lower taxonomic level often become clear when the scale of analysis is enlarged: '[A stepwise comparative] procedure appears to be very virtuous, but in fact is an illusion. The reconstruction will itself be a poorer approximation to the truth if it is confined to a restricted group ... In fact, many phenomena of narrower groups can only be understood historically by outside evidence from within the broader stock' (36). This pronouncement comes quite close to B's notion of TELEORECONSTRUCTION (B 1973)—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 subgroups. B feels in fact that it is often easier to reconstruct at a higher rather than a lower taxonomic level—e.g. Proto-Tibeto-Burman (PTB) is 'easier' to reconstruct than its best-studied subgroup, Proto-Lolo-Burmese (PLB);[26] he claims that the relationship of Japanese to the 'rest of' Austro-Tai (AT) is easier to demonstrate than are the interrelationships of its previously posited branches (Austronesian, Tai-Kadai, and Hmong-Mien).

There is certainly something to be said for a telic approach when one is operating at, say, the macrolevel, where the validity of the language family is obvious to all. But even here it is an illusion to think that the task of reconstruction gets easier when the genetic relationship becomes more remote. It is certainly harder to reconstruct PST than PTB, even though the relationship of Chinese to TB is not in doubt. It is much more difficult to go even one further step up the taxonomic tree from Proto-Tai to Proto-Tai-Kadai, let alone from there back to Austro-Tai or Austro-Tai-cum-Japanese. When a megalo-con-

---

[24] To be fair, we can certainly imagine a protoform-with-respect-to-the-future like *taxicab developing into taxi in Lg. A but into cab in Lg. B; or a protoname like *Elizabeth becoming Lisa in Lg. A but Betty in Lg. B. But it will be all to the good if future linguists can uncover further cognates with shared phonemes (e.g. tekep or Libby).

[25] G uses a false analogy to illustrate this point: 'For example, there is no doubt concerning just which languages belong to the vast Austronesian family, but subgrouping has proved difficult and has not led to any generally accepted result' (59). While it is true that Austronesianists are still not in total agreement on some details of Proto-AN reconstruction, the various subgroups of the family (e.g. Proto-Polynesian) are very well reconstructed indeed, and thousands of valid cognate sets have been established at various taxonomic levels, the vast majority of which illustrate recurrent, regular sound correspondences.

[26] Personal confession, numerous occasions. We know enough about PLB to see how much remains to be explained.

struct is already highly shaky and speculative, the very ease with which it may
be further elaborated should be suspect.

For me the ultimate problem with proposals of super-remote genetic rela-
tionship is the impossibility of choosing among conflicting ones. I have men-
tioned the various theories about the affiliations of Japanese (§3), where G and
B do not see eye to eye. The two also differ on the further genetic affiliations
of Sino-Tibetan, though they both agree it is distinct from Austro-Tai. While
B for once ventures no further affiliations for Sino-Tibetan, G is inclined to
accept Sapir's hypothesis of a relationship between ST and Na-Dene (G 332;
Campbell 1988:593). The eminent Sinologist and historian E. G. Pulleyblank
claims rather that ST's true connections are with Indo-European (1978). G
endorses Schmidt's old idea (1906) of a genetic connection between Austrone-
sian and Austroasiatic (Mon-Khmer plus Munda), and favors adapting the the-
ory to encompass all of B's Austro-Tai as well as Austroasiatic. B once believed
in Schmidt's theory himself (the 'Austric' theory; see Diffloth 1985), but lately
refuses to recognize anything more than an old contact relationship between
Austronesian and Austroasiatic.

CONCLUSION

5. It is no doubt a noble idea to try to find a common genetic origin for all
of the world's languages, just as a common biological origin can be demon-
strated for all extant 'races' of *Homo sapiens*. G appeals to biological genetic
evidence to back up his linguistic genetic claims: 'I would like to emphasize
the fact that my linguistic classification shows an almost exact match with
genetic classification by population biologists and with fossil teeth evidence'
(1989:113). But dental consonant correspondences are much better indicators
of linguistic relationship than are dental fossils. There is no linguistic correlate
of DNA to furnish irrefutable proof of genetic relationship. People of any ge-
netic makeup can speak any language. There seems little reason to believe that
the New World was populated by a discrete series of exactly three waves of
migration, each racially and linguistically uniform. Why not conceive of this
population movement as a millennial percolation of people speaking a variety
of originally unrelated languages, some of which were subjected along the slow
migration routes to prolonged mutual contact and even creolization?

Looking on the positive side, one has to admire G's intellectual vigor and
daring. No one could ever accuse him of timidity or fuddyduddyism. Perhaps
LIA will actually prove to be a boon to Amerindian linguistics because of the
intensity of the counter-research it will engender. In discussing his methodology
G sets up a dichotomy between the traditional comparative method of 'looking
at many forms across only two or a few languages' and his own method of
'looking at few forms across many languages'. Surely what is needed is a com-
bination of the virtues of both approaches: looking at as many well-recorded
and well-analyzed forms from as many languages as possible. It is also a good
idea to organize the data by semantic field as well as by phonological shape.[27]

---

[27] This is the approach, for instance, of Kaufman's *Otomanguean Etymological Dictionary* and
Diffloth's ongoing *Mon-Khmer Etymological Dictionary* project, as well as of my *Sino-Tibetan
Etymological Dictionary and Thesaurus* project at Berkeley.

The computer will be crucial to handle the etymological information explosion of the future, but machines will never be able to do all the work for us. There will still doubtless be room for gut feelings, intuitions, temperamental quirks, and professional rivalries, even in the Computer Age. In making etymological 'judgment calls' there will never be a substitute for hands-on human experience in a given language or language family.

REFERENCES

ADELAAR, WILLEM F. H. 1989. Review of Greenberg 1987. Lingua 78.249–55.
BENEDICT, PAUL K. 1972. Sino-Tibetan: A conspectus. Contributing editor, James A. Matisoff. Cambridge: Cambridge University Press.
——. 1973. Tibeto-Burman tones, with a note on teleoreconstruction. Acta Orientalia 35.127–38.
——. 1975. Austro-Thai language and culture, with a glossary of roots. New Haven: Human Relations Area Files Press.
——. Forthcoming. Japanese/Austro-Tai. [300-page MS version, 1985.] Ann Arbor: Karoma Press.
BURROW, THOMAS, and MURRAY B. EMENEAU. 1961. Dravidian etymological dictionary. Revised Edition. London: Oxford University Press.
CALLAGHAN, CATHERINE A., and WICK MILLER. 1962. Swadesh's Macro-Mixtecan hypothesis and English. Southwest Journal of Anthropology 18.278–85.
CAMPBELL, LYLE. 1988. Review article on Greenberg 1987. Lg. 64.591–615.
CHAFE, WALLACE. 1987. Review of Greenberg 1987. Current Anthropology 28.652–3.
COOKE, JOSEPH R. 1968. Pronominal reference in Thai, Burmese, and Vietnamese. (University of California Publications in Linguistics, 52.) Berkeley & Los Angeles: University of California Press.
DIFFLOTH, GÉRARD. 1985. What happened to Austric? University of Chicago, MS.
DONG WEIGUANG; CAO GUANGQU; and YAN XUEQUN. 1984. Hanyu he Dong-Taiyu de qinyuan guanxi [Genetic relationship between Chinese and Dong-Tai languages]. Computational Analysis of Asian and African Languages (Tokyo) 22.105–21.
GODDARD, IVES. 1986. Sapir's comparative method. New perspectives in language, culture and personality, ed. by William Cowan, Michael K. Foster, & Konrad Koerner, 191–210. Amsterdam & Philadelphia: John Benjamins.
GREENBERG, JOSEPH H. 1987. Language in the Americas. Stanford: Stanford University Press.
——. 1989. Classification of American Indian languages: A reply to Campbell. Lg. 65.107–14.
——. Forthcoming. Indo-European and its closest relatives: the Eurasiatic language family. Stanford: Stanford University Press.
HASHIMOTO, ANNE OI-KAN. 1976. Substratum in Southern Chinese—The Tai connection. Computational Analysis of Asian and African Languages (Tokyo) 6.1–10.
KAUFMAN, TERRENCE S. 1990. Otomanguean etymological dictionary. University of Pittsburgh, MS.
KAWAMOTO, T. 1977/1978. Towards a comparative Japanese-Austronesian. Bulletin of the Nara University of Education 26.23–49 (Part I), 27.1–24 (Part II).
LABBERTON, D. VAN H. 1924. Preliminary results of researches into the original relationship between the Nipponese and the Malay-Polynesian languages. Journal of the Polynesian Society 33.244–80.
LI FANG KUEI. 1976. Sino-Tai. Computational Analysis of Asian and African Languages (Tokyo) 3.39–48.
MARTIN, SAMUEL. 1966. Lexical evidence relating Korean to Japanese. Lg. 43.185–251.
MATISOFF, JAMES A. 1976. Austro-Thai and Sino-Tibetan: An examination of body-part contact relationships. Genetic Relationship, Diffusion, and Typological Similarities

of East and Southeast Asian Languages, ed. by Mantaro J. Hashimoto, 256–89. Tokyo: Japan Society for the Promotion of Science.

——. 1979. Problems and progress in Lolo-Burmese: Quo vadimus? Linguistics of the Tibeto-Burman Area 4.2:11–43.

——. 1982. Proto-languages and proto-Sprachgefühl. Linguistics of the Tibeto-Burman Area 6.2:1–64.

——. 1988. Areal and universal dimensions of grammatization in Lahu. Paper presented at the Symposium on Grammaticalization, University of Oregon. [To appear in Elizabeth Traugott & Bernd Heine, eds., Approaches to grammaticalization, Vol. II. Amsterdam & Philadelphia: John Benjamins.]

MILLER, ROY ANDREW. 1971. Japanese and the other Altaic languages. Chicago: University of Chicago Press.

NISHIDA, TATSUO. 1978. Chibetto-Biruma-go to Nihongo [Japanese and the Tibeto-Burman languages]. Nihongo no keitoo to rekishi [The lineage and history of Japanese] (Kooza Nihongo 12), ed. by Ohno Susumu & Shibata Takeishi, 229–300. Tokyo: Iwanami Shoten.

OHNO, SUSUMU. 1980. Sound correspondences between Tamil and Japanese. (Gakushuin Series of Treatises 8.) Tokyo: Gakushuin Educational Foundation.

PULLEYBLANK, EDWIN G. 1978. Chinese and Indo-European: The phonetic role of the *kan-chih* signs in the Archaic Chinese script and their relation to the alphabet, together with evidence for a genetic relationship between Chinese and Indo-European. Paper presented at the Conference on the Origins of Chinese Civilization, University of California, Berkeley.

RAHDER, JOHANNES. 1956/1959. Etymological vocabulary of Chinese, Japanese, Korean and Ainu. Part I (1956): Monumenta Nipponica (Tokyo), Monograph 16. Parts II and III (1959): privately printed, New Haven.

SCHMIDT, PATER WILHELM. 1906. Die Mon-Khmer-Völker, ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. Archiv für Anthropologie 33.59–109.

SHAFER, ROBERT. 1952. Athabaskan and Sino-Tibetan. IJAL 18.12–19.

SOLOMON, JAMES R. 1974. Japanese and the other Austronesian languages: A survey of the literature. Paper presented at the First International Conference on Austronesian Linguistics, Honolulu.

WEIDERT, ALFONS. 1987. Tibeto-Burman tonology. Amsterdam & Philadelphia: John Benjamins.

WULFF, KURT. 1934. Chinesisch und Tai. Det Kgl. Danske Videnskabernes Selskab, Historisk-filologiske Meddelelser 20.3.

YAN XUEQUN. 1983. Miao-Yao Hanyu guanxici de cengci [Strata of related words in Miao, Yao, and Chinese]. Paper presented at the 16th International Conference on Sino-Tibetan Languages and Linguistics, University of Washington.

ZHAO YANSUN. Forthcoming. [Bai-Chinese Dictionary.] Edited by Grace Wiersma.

Department of Linguistics   [Received 27 November 1989.]
University of California
Berkeley, CA 94720