

# 10

## Taxonomy, Typology and Historical Linguistics

Merritt Ruhlen  
*Stanford University*

### 1. Introduction

The past decade has witnessed a renewed interest in historical linguistics, as the various controversies surrounding Amerind, Nostratic, and even broader proposed taxa well attest. Yet this renewed interest seems to have revealed as much the current state of confusion within historical linguistics as the validity of any of the newly proposed families. I will argue here that the comparative method was misunderstood by historical linguists in the twentieth century, with the result that the discovery of new genetic relationships among languages effectively ground to a halt — with the significant exceptions of the work of Joseph Greenberg and the Nostraticists. What is equally distressing is that the borders between three distinct fields — taxonomy, typology, and historical linguistics — have become blurred. Each of these fields has its own goals and its own methodology, and they are not the same. This in no way implies that these fields are completely disconnected from one another. Certainly Greenberg's enormous knowledge of diachronic typology informed his classification of Eurasiatic languages in many ways, most spectacularly in the explanation of the origin of the

Indo-European ablaut system and its historical connections with the vowel harmony systems of Uralic and Altaic (Greenberg, 2000). In the same way, his knowledge of historical linguistics is utilized from the very first steps in taxonomy (Greenberg, 1995), allowing him to recognize the most obvious etymologies and to weed out some spurious ones. What characterizes his work is that he used all three fields in an appropriate manner, and did not confuse the goals and results of these three different fields.

## 2. Taxonomy and Historical Linguistics

The source of a great deal of the current controversy and confusion in historical linguistics resides in the fact that taxonomy<sup>1</sup> failed to develop in the twentieth century beyond the obvious, and in some instances even regressed. One of the more egregious examples of regression is Johanna Nichols' (1992: 4) rejection of the Altaic family:

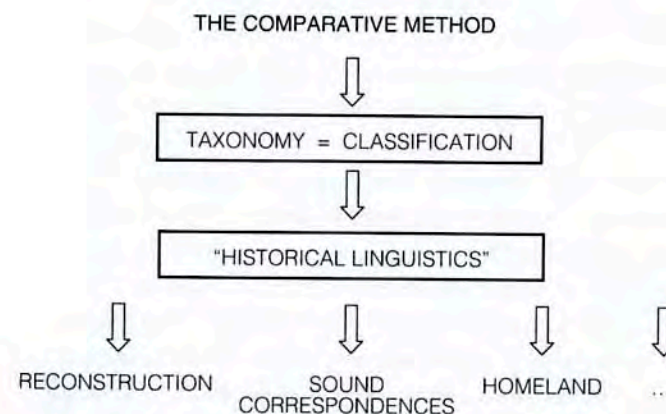
Three language families of central Eurasia . . . share striking similarities in morphosyntactic structure and pronominal roots: Turkic, Mongolian, and Tungusic. For a long time it was assumed that these three families were related as branches of a superstock called Altaic . . . . When the cognates proved not to be valid, Altaic was abandoned, and the received view now is that Turkic, Mongolian, and Tungusic are unrelated.

It is unlikely, however, that this 'received view' would be accepted by Roy Andrew Miller (1971, 1991a,b,c), Sergei Starostin (1991), Anna Dybo, Oleg Mudrak (Dybo, Mudrak, and Starostin, 2003), or Greenberg (2000–02) and indeed the Altaic family is in no more doubt today than it was a century ago.

<sup>1</sup> Also called classification, multilateral comparison, and mass comparison. All four terms are synonymous.

If one consults virtually any of the standard textbooks on historical linguistics the subject of taxonomy is not even mentioned and the precise means by which one discovers new language families is either not presented, or is presented in a completely fictitious manner in which reconstruction and sound correspondences are alleged to be the proof of language families. In fact, the fundamental error of the past century was that the comparative method came to mean, in linguistics, the reconstruction of a proto-language using regular sound correspondences. For example, in the index to Bynon (1977: 305) we find "comparative method, the *see* reconstruction, phonological." In reality the comparative method in linguistics, biology, or any other field consists of essentially two stages. The first is taxonomy and the second is what is commonly called "historical linguistics," as shown in Figure 1.

Figure 1. The comparative method



These two stages are to a very great degree independent of one another and it is taxonomy, not historical linguistics, that defines language families at all levels. Taxonomy provides the wherewithal for the pursuits of most historical linguists: (1) the reconstruction of the proto-language, (2) the discovery of sound correspondences



among the constituent languages (or families), (3) the subgrouping of the family, (4) the location of the ancestral homeland, etc. In Figure 1 I have listed reconstruction and sound correspondences as independent parameters for the simple reason that reconstruction can be carried out in fields such as biology where sound correspondences, or anything analogous to them, are absent. It should also be noted that in biology reconstruction is in no way identified with the comparative method as it is in linguistics. In fact, in discussions of the comparative method in biology reconstruction is scarcely mentioned, and no biologist has ever demanded that Proto-Mammal be reconstructed, along with all of the intermediate stages leading from Proto-Mammal to all modern species of mammal, before he believes that mammals are a valid biological taxon.

It has been alleged by a number of scholars that Greenberg has substituted for the comparative method an entirely different method for the investigation of linguistic prehistory. According to the traditional view “the comparative method does not apply at time depths much greater than about 8,000 years” (Nichols, 1992: 2). Greenberg’s methods supposedly begin at this cut-off point and produce families that are qualitatively different from those produced by the standard comparative method. Mark Durie and Malcolm Ross (1996: 5, 9) claim that “multilateral comparison is not a variant of the classical comparative method of historical linguistics. . . . Multilateral comparison . . . bears only the most superficial resemblance to the comparative method.” Goddard and Campbell (1994: 195) believe “the differences between Greenberg’s word-comparison approach and the standard historical-linguistic method are so vast that rational discussion between their respective proponents seems almost impossible.” Hans Hock and Brian Joseph (1996: 487) allege that “the American linguist Joseph Greenberg and some associates of his have claimed that long-distance relationships can be established more effectively — and more easily — by employing an approach totally different from the traditional methods. This is an approach of lexical ‘mass comparison’ or ‘multilateral comparison.’” And according to Nichols (1990: 477), “Greenberg (1987) makes clear that he believes such groupings [as

Altaic, Hokan, and Amerind] cannot be reached by the standard comparative method; a wholly different method, *mass comparison*, is required.” In reality, no such claim is made in Greenberg (1987), nor anywhere else in his writings, for the simple reason that in his view the comparative method applies in the same way at all levels of taxonomy, from the lowest to the highest.

This confusion between the different stages of taxonomy and historical linguistics is apparent in the following quote from Bynon (1977: 271–72): “The use of basic vocabulary comparison not simply as a *preliminary* to reconstruction but as a *substitute* for it is more controversial . . . . It is clear that, as far as the historical linguist is concerned, [mass comparison] can in no way serve as a substitute for reconstruction.” But Greenberg never claimed that taxonomy is a substitute for reconstruction. Taxonomy and reconstruction are two separate and distinct enterprises. Taxonomy identifies families at all levels; reconstruction seeks to reconstruct the proto-language of a family that has *already been identified* by taxonomy. It is hard to see how one could even begin to reconstruct a proto-language of a family that hadn’t yet been discovered, much less that such a reconstruction would then somehow ‘prove’ the validity of that language family. A similar confusion of taxonomy with reconstruction is seen in Terrence Kaufman’s (1990: 23) claim that “a temporal ceiling of 7000 to 8000 years is inherent in the methods of comparative linguistic *reconstruction*. We can recover *genetic relationships* that are that old, but probably no earlier than that” (italics added). A third example of the confusion of reconstruction and classification appears in a recent textbook (Fox, 1995: 236): “One of the most controversial developments in the whole field of *reconstruction* in recent years has been the publication of Joseph Greenberg’s *classification* of the native languages of the Americas” (italics added). Many additional quotes could be adduced, but these three more than suffice to show that the distinct notions of genetic relationship and reconstruction have become almost synonymous in the minds of many linguists. In reality, genetic relationships are *properties* of classifications; they are not consequences of reconstruction.

Sometimes a distinct taxonomic stage is recognized as preceding the later stages that I have called historical linguistics. Yet even here the precise nature of taxonomy seems poorly understood. For example, Durie and Ross (1996: 6–7) have recently characterized the comparative method as consisting of seven stages, the first two of which correspond to what I have called taxonomy, and the last five to historical linguistics. The first two stages are:

1. Determine on the strength of diagnostic evidence that a set of languages are genetically related, that is, that they constitute a ‘family.’
2. Collect putative cognate sets for the family (both morphological paradigms and lexical items).

What is peculiar here is that the two initial steps are given in reverse order. One first uses some mysterious ‘diagnostic evidence’ to identify a language family and then one goes out and actually looks for evidence (grammatical and lexical) to support the validity of the family. But putative cognate sets *are* the diagnostic evidence for any family. It is the recognition of grammatical and lexical resemblances *in both form and meaning* that leads to the supposition that certain languages (or language families) are genetically related. The reason that such putative cognate sets are the basis for detecting linguistic relationships is so obvious that it is often seemingly overlooked. The basis of genetic classification — and hence linguistic relationships — is, quite simply, the arbitrary nature of the sound/meaning relationship in human language. Since any meaning can be represented by any sequence of sounds, there are hundreds, if not thousands, of possible phonetic representations for each meaning in each language. If, then, a certain set of languages has the same, or similar, phonetic representation for a word, one assumes the languages may well be related. If further consideration of this set of languages shows additional similar words, to the exclusion of other surrounding languages, the hypothesis of genetic relationship becomes virtually certain. Any language may accidentally resemble another language once, but if the resemblances continually appear in the *same* set of languages, and not elsewhere, they are hardly likely to be accidental.

In addition to chance, there are three other possible explanations for resemblant words: sound symbolism, borrowing, and common origin. Sound symbolic words are quite exceptional, precisely because they violate the arbitrary sound/meaning relationship, and are not used in classification, though they are quite correctly reconstructed for proto-languages since all languages do have sound symbolic words. Borrowings can usually be recognized by well-known linguistic techniques, such as reliance on basic vocabulary (pronouns, body parts) and outgroup comparison with other related languages. If the languages concerned are never known to have been in contact — or if the languages concerned cover entire continents — then borrowing is quite improbable.

In recent years historical linguists have come to regard common origin, that is, an evolutionary explanation for linguistic similarities, as the explanation of last resort, when in fact it is, as Vincent Sarich (1994) pointed out, the default explanation. Instead of recognizing the simple basis of genetic classification, and thus linguistic relationships, twentieth-century historical linguists put forth increasingly rigorous demands, generally involving reconstruction with regular sound correspondences, before genetic relationships will be acknowledged, demands in fact so rigorous that they could never be satisfied. According to Calvert Watkins (1990: 292–95),

a genetic linguistic relationship is first assumed, or hypothesized, by inspection or whatever. At that point must begin the careful and above all systematic comparison, which will lead, if the hypothesis or supposition of genetic relationship is correct, to the reconstruction of the linguistic history of the languages concerned, including the discovery of the attendant sound laws, which are a part of that history. . . . If I believe in Indo-European, Algonquian, or Austronesian, it is because scholars have done the necessary systematic explanation and produced the requisite historical results. If I do not believe in an Amerind, Eurasiatic, or Nostratic, it is because scholars have so far neither done the one nor produced the other. To spell it out: because



scholars have neither done the necessary systematic explanation, nor produced the requisite historical results. And there is no other way.

But there is another way. It is the way that was used by the founders of comparative linguistics in the nineteenth century. And it is the same method advocated, and utilized, by Greenberg in all his works. This 'other way' is known as taxonomy, classification, mass comparison, or multilateral comparison. To demand, as does Watkins, that one must reconstruct the entire proto-language and then explain with regular sound correspondences exactly how every word in every descendant language evolved to its present form is clearly far more than was ever demanded of the very families that Watkins cites approvingly: Indo-European, Algonquian, and Austronesian. All three of these families were recognized early on simply by the specific, and distinct, grammatical and lexical morphemes that characterize each. Reconstruction, regular sound correspondences, and a complete explanation of all linguistic prehistory was never demanded and in fact such concepts as reconstruction and the regularity of sound change first appear only in the second half of the nineteenth century, long after these particular families were accepted by everyone. The recent claim by Nichols (1996: 46) that "the philologist of [Sir William] Jones' time had been trained . . . in the principles of comparative method and reconstruction" is so wildly anachronistic as to defy explanation.

The goal of an historical linguist should not be to demand an explanation for everything before he believes anything, as Watkins implies. Rather a scientist should attempt to explain non-random phenomena, for example, the prevalence of the N/M 'I/thou' pronominal pattern in the Americas, and a different pronominal pattern, M/T 'I/thou' in northern Eurasia. It is all well and good for Watkins to demur on the Amerind and Eurasiatic hypotheses, but if common origin is not responsible for these different pronominal patterns, what is?

The requirement of reconstruction with regular sound correspondences appears to have been an innovation of twentieth-century historical linguists. It is not found, so far as I

know, in any of the works of nineteenth-century pioneers of comparative grammar such as Karl Brugmann or Berthold Delbrück. In the late nineteenth century Indo-European was being reconstructed; no scholar thought he was 'proving' Indo-European, much less discovering it. The family was accepted by everyone, and that was why they were trying to reconstruct it. What then did the pioneers of comparative Indo-European take as the basis of genetic relationship, if not reconstruction? The following quote from Delbrück is instructive:

My starting point is that specific result of comparative linguistics that is not in doubt and cannot be in doubt. It was proved by Bopp and others that the so-called Indo-European languages were related. The proof was produced by juxtaposing words and forms of similar meaning. When one considers that in these languages the formation of the inflectional forms of the verb, noun, and pronoun agrees in essentials and likewise that an extraordinary number of inflected words agree in their lexical parts, the assumption of chance agreement must appear absurd (Delbrück, 1880: 121–2).

Greenberg has been accused of having attempted to substitute taxonomy for reconstruction, but what really happened in the twentieth century was that historical linguists sought to substitute reconstruction for taxonomy, thus confusing the goals of historical linguistics with the requirements of genetic classification.

### 3. Pronouns

It is not by accident that pronouns have figured as one of the major foci of taxonomic controversies in the twentieth century. As Dolgopolsky showed in 1964, the first- and second-person pronouns are the first and third most stable meanings in language (the numeral 'two' is second). The past decade witnessed an endless controversy over the alleged Amerind N/M pattern, and only slightly



less controversy over the Eurasiatic M/T pattern.<sup>2</sup> There have been two camps. The first camp sees both patterns as survivals of two different languages, Proto-Amerind in the first case and Proto-Eurasiatic in the second. The second camp has two subgroups. The first subgroup claims that the alleged patterns are specious and that both patterns occur in both the Americas and the Old World; the second subgroup admits the reality of the two patterns, but attempts to give a non-evolutionary explanation in terms of sound symbolism, language universals, diffusion, etc.

We should begin by noting that both the Eurasiatic and Amerind pronominal patterns were clearly recognized at the start of the twentieth century by, among others, Alfredo Trombetti (1905), who abundantly documented the Amerind pattern throughout the Americas in an appendix to his book and concluded:

As can be seen, from the most northern regions of the Americas the pronouns NI 'I' and M 'thou' reach all the way to the southern tip of the New World, to Tierra del Fuego. Although this sketch is far from complete, due to the insufficient materials at our disposal, it is certainly sufficient to give an idea of the broad distribution of these most ancient and essential elements (p. 208).

But Trombetti knew equally well that this American pattern was absent in northern Eurasia, where a totally different pattern, M/T, predominated, and he lamented in his book that "it is clear that in and of itself the comparison of Finno-Ugric *me* 'I,' *te* 'you' with Indo-European *me-* and *te-* is worth just as much as any comparison one might make between the corresponding pronominal forms in the Indo-European languages. The only difference is that the common origin of the Indo-European languages is accepted, while the

<sup>2</sup> The Eurasiatic family consists of Indo-European, Uralic, Altaic, Korean-Japanese-Ainu, Gilyak, Chukchi-Kamchatkan, and Eskimo-Aleut; see Greenberg (2000-02).

connection between Indo-European and Finno-Ugric is denied" (p. 44). Antoine Meillet also was aware of the Eurasian M/T pattern, but he proposed a universal explanation rather than a genetic one: "It goes without saying that, in order to establish linguistic affinity, one must ignore everything that can be explained by general conditions common to all languages. For example, pronouns must be short, clearly made up of consonants that are easy to pronounce, and usually without consonant clusters. It is for this reason that pronouns are similar almost everywhere, without this fact implying a common origin" (Meillet, 1965: 89).

Apparently unaware of Trombetti's appendix on the Amerind pronominal pattern, Edward Sapir a decade later noted the presence of both first-person N and second-person M throughout the Americas and wrote, in a personal letter, "how in the Hell are you going to explain general American *n-* 'I' except genetically?" (quoted in Greenberg, 1987). Franz Boas was also aware of the widespread American pattern, but opposed the genetic explanation given by Trombetti and Sapir: "the frequent occurrence of similar sounds for expressing related ideas (like the personal pronouns) may be due to obscure psychological causes rather than to genetic relationship" (quoted in Haas, 1966). It would thus seem that at least the reality of the pattern — and its virtual restriction to the Americas — was beyond doubt by the beginning of the twentieth century.

Such an assumption would, however, be incorrect, for during the final decade of the twentieth century there was a sharp debate, not just on the proper explanation of the Amerind pattern — genetic or non-genetic — but indeed on its very existence. According to Lyle Campbell (1994a: 47), "the *n/m* ['I/you'] pattern is not nearly as common in the Americas as Greenberg claimed . . . [and] his supposed *m/t* ['I/you'] pattern for his Eurasiatic languages is also found abundantly in the Americas (despite his and Ruhlen's assertions to the contrary)." Campbell also claims that "several Amerind groups exhibit pronoun forms (*m/t* ['I/you']) that Greenberg attributes to Europe and Northern Asia" and "the *n* 'first person' / *m* 'second person' is by no means unique to, diagnostic of, or ubiquitous in American Indian languages" (Campbell,



1994b: 3, 9). Campbell's denial of the reality of both the Eurasiatic and Amerind pronominal patterns is by no means idiosyncratic. Many other scholars have endorsed this view. According to Nichols (1992: 261), "the root consonantism of personal pronouns turns out to have symbolic properties comparable, in both their universality and their basic structural design, to those of "mama"—"papa" vocabulary. . . . Specifically, personal pronoun systems the world over are symbolically identified by a high frequency of nasals in their roots."

In what can only be considered a comical coda to the pronoun controversy the final debate of the twentieth century was between two scholars, Nichols and Campbell, who had previously been on the same side, vigorously opposing a genetic explanation for the Amerind pattern and, in fact, vigorously opposing Amerind as well. In 1996 Nichols published a paper (with David Peterson) recognizing that her previous universal explanations were incorrect and that scholars from Trombetti to Greenberg, who identified different pronominal patterns in different areas of the world, were correct: "the *n:m* paradigm . . . clearly . . . cannot be due to universals or random chance. . . . The *n:m* pronominal system is exceedingly rare outside of the Americas and very common in a geographically limited, though large, part of the Americas" (pp. 336–337). What is left unexplained by Nichols and Peterson is how the methodology of 'population typology' can arrive at precisely opposite conclusions on the basis of the *same* language sample. Again following Greenberg (1987: 54), they reject borrowing as a source of the pronominal similarities, pointing out that "pronouns are almost always inherited" (p. 337).

While it is welcome that Nichols and Peterson (henceforth, N&P) have now independently confirmed Greenberg's conclusion that the Amerind pattern is essentially an American phenomenon and cannot be explained by universals, chance, or borrowing, it is distressing that they falsely claim that theirs is the first proper demonstration. Criticizing both Campbell and me, they claim that "both sides cite only the evidence supporting their claims, and neither cites enough of that positive evidence to convince the reader

of the distribution of the *n:m* pronominal system in Amerind or elsewhere; neither side offers a proper survey than can capture evidence, both positive and negative, without bias so that the field can assess the distribution and status of this pronominal system" (p. 337). In fact, this was precisely what I had done in the two papers that N&P cite (Ruhlen, 1994b, 1995a). In the first of these articles I attempted to survey all the pronominal patterns that had been posited for all the world's language families. For families such as Indo-European or Uralic this merely entailed listing the reconstructions for those families. For families that have not been reconstructed, such as the sub-Saharan families, I listed those pronouns that have been identified by specialists in these families, though without reconstructions. The evidence given was, therefore, neither "positive" nor "negative." In the second article, using Greenberg's 21 Amerindian notebooks, I surveyed the existence of both the N/M and M/T patterns in the Americas. Specifically, I looked for languages in the Americas that had *both* first-person singular N and second-person singular M; at the same time I also looked for languages that had *both* first-person singular M and second-person singular T. The results of this search are shown in Tables 1 and 2.<sup>3</sup>

In addition to the enormous breadth of the distribution of the Amerind pattern, there is a significant depth to the Amerind pattern as well. As can be seen, the Amerind pattern has been reconstructed for many Amerind subgroups, sometimes at great time depths (e.g., Proto-Algic, Proto-Hokan, Proto-Penutian, Proto-Uto-Aztecan, Proto-Tanoan, Proto-Quechuan, Proto-Chibchan, Proto-Aruak, Proto-Guahiban), while the Eurasiatic pattern has never been reconstructed by anyone for any American family, no matter how shallow, except for Proto-Eskimo-Aleut, the easternmost branch of

<sup>3</sup> Since the original article was published I have developed an independent database containing reconstructions from all levels of Amerind (Ruhlen, 2002) and the results of this database have been incorporated in Table 1. Table 2 remains unchanged because no one, to my knowledge, has ever reconstructed M/T for any Amerind group.



Eurasiatic. These results speak for themselves. Moreover, were I to cite languages that have *either* first-person N *or* second-person M, we would find the number of languages cited in Table 1 would increase dramatically, while citing languages with *either* first-person M *or* second-person T would increase Table 2 only modestly since *neither* pronoun is common in the Americas.

Although N&P now agree with Greenberg on the facts of the distribution, as well as on the illegitimacy of explaining these facts in terms of universals, chance, or borrowing, they do not agree with Greenberg that the explanation is genetic: “If Amerind were a genetic reality and the *n:m* paradigm a marker of it, then the marker should have a fairly even distribution over all of Amerind and should be found only there” (p. 367). Yet N&P concede that “the *n:m* paradigm is attested in all six branches of Greenberg’s Amerind” (368), as Table 1 attests. They object apparently that the paradigm is not preserved uniformly in lower-level subgroups, but there is really no such requirement in historical linguistics and there should be no such expectation.

N&P offer two reasons that the *n:m* paradigm is not a genetic marker of Amerind. First, this paradigm in the Americas must be older than the temporal limits of the comparative method and therefore cannot be considered genetic evidence. Secondly, they point out two examples of the *n:m* paradigm in the Old World, the Vanimo<sup>4</sup> language (spoken on the coast of northern New Guinea) and Mongolian, and on the basis of these two languages they project the *n:m* paradigm back to an even earlier historical connection in Asia — the “Pacific Rim distribution” — though they are unable to say whether this historical connection was due to common origin, borrowing, or something else.

Table 1  
Distribution of N ‘I’ – M ‘thou’ in the Americas.

ALMOSAN: Proto-Algic *\*-Vn/\*-Vm*, Kutenai *-na:p/-m*; PENUTIAN: Proto-Penutian *\*n/\*m-*, Tsimshian *n/m-*, Chinook *n/n-*, Proto-Coos-Takelma *\*n/\*ma*, Takelma *-n/ma*, Proto-Plateau-Penutian *\*ni/\*mis*, Proto-Sahaptian *\*?i:n/\*?i:m*, Nez Perce *?i:n/?i:m*, Klamath *ni-/mi-*, Proto-California Penutian *\*ni/\*mVn*, Wintu *ni-/mi-*, Colouse *nat/mit*, Patwin *na-/mi-*, Proto-Yokuts *\*na-/\*ma-?*, Proto-Maiduan *\*niki/\*min*, Maidu *ni/mi*, Nisenan *ni/mi*, Proto-Miwok-Costanoan *\*ka-na/\*mi*, Tunica *-ni/ma*, Huave *-na-/me-*; HOKAN: Proto-Hokan *\*n<sup>s</sup>a/\*ma*, Chimariko *no-/mam-*, Karok *na/im*, Arra-arra *na/im*, Pehtsik *naah/cehm*, Washo *le (< \*na)/mi*, Esselen *niš-/miš-*, Proto-Yuman *\*ñ-/\*m-*, Yuma *nyep/mañ*, Mohave *inyeč/manč*, Walapai *añ/ma*, Havasupai *inyal/ma-a*, Yavapai *nya-a/ma-a*, Diegueño *?enyaal/maa*, Chontal *ni/mi*, Coahuilteco *na/mai*, Comecrudo *na/emnã*, Karankawa *n/im*, Cotoname *na/men*; CENTRAL AMERIND: Proto-Aztec-Tanoan *\*ne?/\*?eme*, Proto-Tanoan *\*nō/\*?ēm*, Kiowa *nā/am*, Jemez *ne/ūmiš*, Proto-Uto-Aztecan *\*n/\*m*, Kawaiisu *ni/?i/imi*, Utah *ne/yim*, Opathe *?ina-po/?emé?e*, Yaqui *?inapo/?eme?e*, Tarahumara *ni-hé/?yemi*, Papago *-ñ/-m*, Hopi *nu/?uma*, Nahuatl *no-/mo-*, Pipil *-neč/-mit<sup>s</sup>*; CHIBCHAN: Proto-Chibchan *\*na-sV/\*mue-ya*, Cogui *nós/má*, Proto-Aruak *\*na/\*ma*, Ica *nən/ma*, Chimila *náari/ámma*, Bribri *ñō/ma*, Rama *na/ma*, Miskito *yan (? < \*ñan)/man*, Ulua *yan/man*, Sumu *yan/man*, Guamaca *nerra/ma*, Lenca *una/amna*; ANDEAN: Proto-Quechuan *\*nuqa/\*qam*, Jaqaru *na/huma*, Aymara *naya/huma*, Mapudungu *ta-ñi/ta-mi*; EQUATORIAL: Proto-Guahiban *xáni/xámi*, Mocoche *an-/ma*, Pakaasnovos *na/?wum*, Achuar *wina/amin*, Jitnu *kan/kam*, Cuiba *xan/xam*, Guayabero *xan/xam*, Itene *ana-/ma-*; MACRO-PANOAN: Moseten *ñu/mi*, Nocten *no-/em*, Pacaguara *no-/mi-*, Chacobo *no-/mina*, Arazaire *nocna/mina*; MACRO-GE: Delbergia *nū/ma*, Tibagi *in/ama*, Catarina *enha/ahama*, Kaingang *?iñ/?ã*.

Table 2  
Distribution of M ‘I’ – T ‘thou’ in the Americas.

ESKIMO-ALEUT: Proto-Eskimo *\*-ma* ‘I, my,’ Sirinik *məŋa* ‘I,’ elsewhere: *uvanga* ‘I,’ *\*-t (> -n* in Western dialects) ‘thy,’ *\*-ti-k* ‘your dual,’ *\*-ti-t* ‘your plural’; AMERIND: Siouan: Mandan *mī/da*, Hidatsa *ma/da*, Paezan: Millcayac *mioiñ/təz*, Ge: Coroado *make/teke*.

<sup>4</sup> Campbell (1997: 339) incorrectly identifies this language as Austronesian. It belongs rather to the Indo-Pacific family (Greenberg, 1971: 822).



Let us dispense first with the Mongolian example. N&P consider Mongolian *nam-* ‘me,’ *čam* ‘thee’ to be an example of the “*n*:-*Vm* paradigm,” itself a variant of the *n:m* paradigm. The problem here is that the *-m* of *ča-m* is an accusative marker, not a second-person pronoun; the second-person marker is *ča-*, which derives from Classical Mongolian *čima-*. This stem contains *či-* ‘thou,’ itself deriving from Proto-Altaic \**ti-*, as comparison with the plural pronoun *ta* ‘you’ indicates (Illich-Svitych, 1976: 49). While most linguists would consider the comparison of an accusative ending with a second-person marker a rather serious error, apparently N&P do not, for in response to a similar criticism by Campbell (1997: 344) their reply was that “in sampling we do not purport to give accurate descriptions of the histories of language families; we want to make meaningful comparisons of the frequency of *n:m* systems in the Americas vs. elsewhere” (p. 609). Most linguists, however, will have a difficult time considering the comparison of an accusative marker with a pronoun a “meaningful comparison,” when the actual history of the form is well known.

As for the sole New Guinea example, Campbell (1997: 346) is certainly correct that “chance congruence between parts of New Guinea and parts of America is a much more plausible account than that there is a mysterious historical connection between just these two regions that defies both time and space and is beyond standard notions of language change associated with the comparative method.”

If one considers the historical implications of N&P’s Pacific Rim theory, one immediately encounters contradictions. For example, N&P claim that “the languages of eastern North America and especially those of eastern South America presumably descend from earlier colonizations” (p. 369), that is, earlier than the later migration that spread the N/M pattern throughout western North and South America. Accordingly, N&P consider the Algonquian family to be the result of an earlier migration because it lacks the crucial N/M pronoun system. But Algonquian’s closest kin — Wiyot and Yurok — are (or were) spoken on the California coast and Paul Proulx (1985) has reconstructed the N/M paradigm for Proto-Algic

(= Algonquian, Wiyot, and Yurok) and Algonquian itself has preserved first-person N. Since Proto-Algic possessed the N/M pattern, as Sapir (1913) thought, this simply means that second-person M was lost in Algonquian.<sup>5</sup> But how can the loss of a single pronoun and a movement to the eastern seaboard overturn this close genetic relationship with Wiyot and Yurok and transform Algonquian into an earlier migration to the Americas? Of course as soon as one looks at other traits one finds abundant resemblances between Algonquian and all the other Amerind groups in western North America. Algonquian is just a normal Amerind group that happened to lose the second-person Amerind pronoun M.

Campbell (1997) makes a number of astute observations and appropriate criticisms of N&P’s article, including pointing out that the alleged absence of the N/M pattern in eastern North and South America has been greatly exaggerated by N&P on the basis of a poorly chosen sample. But what really disturbs Campbell is that he realizes N&P have inadvertently walked into a trap of their own making. Having eliminated borrowing, accident, and universals as possible explanations N&P have left themselves (and Campbell) only one remaining possible explanation, which Campbell notes with alarm: “In denying borrowing of pronoun patterns, N&P in effect rule out ‘areal affinity’ [diffusion] and thus limit, perhaps unwittingly, the interpretation of their ‘single historical development’ to only one possible explanation: genetic relationship, inheritance from a common ancestor” (p. 341). And N&P’s vague description of this mysterious unknown historical process that led to the N/M distribution, “a single historical development of *some sort . . . some kind* of shared history” (p. 337, italics added), is not likely to satisfy anyone. Nor is their final response to Campbell on this very question: “Something happened. We do not and cannot know just what happened, but this does not preclude establishing when and where” (p. 613). What happened is the simple genetic

<sup>5</sup> For the origin of the second-person pronoun that replaced M in Algonquian, see Greenberg (1987: 287).



explanation — a single population entered the Americas with the N/M pronoun pattern, spread rapidly throughout both North and South America around 11,000 years ago according to the archaeological record (Klein, 1999) and left in its wake traces of the original N/M pattern and numerous other diagnostic Amerind traits.

#### 4. Lexical Evidence

If the evidentiary value of pronouns has been swept under the rug in recent decades, the value of lexical evidence has come to be entirely discounted by many scholars, who often refer to it disparagingly as the ‘laundry list’ approach. According to Goddard and Campbell (1994: 195), “Greenberg’s classification [of American Indian languages] is a codification of his judgements of inspectional similarity and is thus, in principle, ahistorical.” Robert Rankin (1992: 330) proclaims that “the days are gone when ‘word-list linguistics’ could be profitably practiced by American linguists and anthropologists.” The reality is just the opposite. Scholars such as Rankin, Goddard, and Campbell have become so specialized in a narrow subfield, preoccupied with small obvious language families such as Algonquian, Siouan, or Mayan, that they are unaware what roots are widespread in the Americas, and which are not. In other words, they have not even reached the word-comparison stage, much less surpassed it.

A single lexical element from the Amerind family illustrates this point. As shown in Ruhlen (1994c), there is a plethora of forms throughout North and South America with the shape *tVnV* and the meaning ‘child, son, daughter’ or the like. Furthermore, a careful analysis of hundreds of such forms suggests that the first vowel of this root was originally correlated with the gender of the child, with *i* indicating masculine gender, *u* feminine, and *a* indeterminate sex. Thus Proto-Amerind must have had a morphologically complex root \**t’ina* ‘son, brother,’ \**t’una* ‘daughter, sister,’ and \**t’ana* ‘child, sibling.’ No extant Amerind language preserves all three grades of

this root intact, but a number do preserve two (e.g., Tiquie *ten* ‘son,’ *ton* ‘daughter’), and even more preserve one. All three grades are, however, preserved elsewhere, for example in the Tucano numeral for ‘one’: *nik-e* ‘one (masc.),’ *nik-o* ‘one (fem.),’ *nik-a* ‘one (indet.)’ (Giacone, 1949). In this case all three grades are retained in what is the general Amerind word for ‘one’ (Ruhlen, 1995b). From the hundreds of examples given in Ruhlen (1994c), Table 3 gives one example of each grade of the root for each of the 13 Amerind subgroups. For Greenberg’s critics there is *no* historical connection between any of these forms.

In addition to the multitude of forms where the vowel still indicates the appropriate gender, there are many forms that are clearly cognate with these forms, but in which the gender appears anomalous from the perspective of the Proto-Amerind pattern. There are numerous well-known typological developments that can lead to this situation. One is when the indeterminate gender becomes specialized as either masculine (e.g., Yuchi *tane* ‘brother’) or feminine (e.g., Proto-Siouan \**i-thā-ki* ‘man’s sister’). Another example is Proto-Algonquian \**ne-tāna* ‘my daughter,’ but already in 1923 Sapir had recognized that “Proto-Algonquian \**-tan-* must be presumed to have originally meant ‘child’ . . . and to have become specialized in its significance either to ‘son’ (Wiyot) or ‘daughter’ (Algonkin proper), while in Yurok its close relative *-ta-tʰ* [‘child’] preserved a more primary genetic significance” (Sapir, 1923: 41). Furthermore, this morphologically complex item occurs with a whole variety of Amerind suffixes and prefixes, as shown in Ruhlen (1994c).

Clearly it is much more likely that all these *t’inalt’analt’una* forms are related historically than that they are not. And that is the question that must be kept in mind. One should not demand that Proto-Amerind be reconstructed with regular sound correspondences; rather one should ask what is the most probable explanation for the presence of the Amerind pronominal pattern N/M in the Americas (and its absence elsewhere) and the presence of a morphologically-complex root \**tVnV*, exhibiting a distinctive gender ablaut system (not found elsewhere in the world), all *in the*



Table 3. Proto-Amerind \*T'INA/\*T'ANA/\*T'UNA "son/child/daughter"

PROTO-AMERIND	*T'ANA "CHILD, SIBLING"	*T'INA "SON, BROTHER, BOY"	*T'UNA "DAUGHTER, SISTER, GIRL"
Almosan	t'an'a "child" (Nootka)	t'in "young man" (Yurok)	tune "niece" (Coeur d'Alene)
Keresiouan	tane "brother" (Yuchi)	-t'sin "male boy" (Mohawk)	-t'aona "sister" (Keres)
Penutian	t'ána-t "grandchild" (Totanae)	pné-t'in "my elder brother" (Molala)	túne- "daughter" (Cent. Sierra Miwok)
Hokan	t'an-pam "child" (Coahuilteco)	t'iní-si "child, son, daughter" (Yana)	a-t'on "younger sister" (Salinan)
Central Amerind	*tana "daughter, son" (Uto-Aztecan)	?límínó "brother" (Cuicatec)	-t'ut'ina "older sister" (Taos)
Chibchan	tuk-tan "child, boy" (Miskito)	sin "brother" (Changuena)	tuntu-rusko "younger sister" (Lenca)
Paezan	dani- "mother's sister" (Warrau)	tzhung "son" (Milkayac)	t'ub-ki "sister" (Cayapa)
Andean	tayna "first-born child" (Aymara)	den "brother" (Tehuelche)	thau "sister" (Tehuelche)
Macro-Tucanoan	tani-mai "younger sister" (Masaca)	ten "son" (Tiquie)	ton "daughter" (Tiquie)
Equatorial	ta?in "child" (Urubu-Kaapor)	tin-gwa "son, boy" (Mocochi)	a-tune-sas "girl" (Morotoko)
Macro-Carib	tane "my son" (Pavishana)	dénu "male child" (Yagua)	-tona "sister" (Nontuya)
Macro-Panoan	tavin "grandchild" (Lengua)	u-tse-kwa "grandchild" (Tacana)	-tóna "younger sister" (Tacana)
Macro-Ge	tog-tan "girl" (Tibagi)	éina "older brother" (Guato)	a-ton-ká "younger sister" (Piokobyé)

same set of languages? Only common origin is a reasonable explanation for these and the many other grammatical and lexical items that characterize the Amerind family.

## 5. Typology

If the confusion between taxonomy and historical linguistics was pervasive throughout the twentieth century, the fundamental distinction between typological classification and genetic classification — the former based on historically-independent structural traits, the latter on historically-related genetic traits (i.e., those involving both sound and meaning) — seemed securely established by Greenberg's African classification in which, for the first time, typological traits were eliminated from consideration. Greenberg's elimination of typological traits from genetic classification has been hailed by many as one of the primary achievements of this African classification (Dimmendaal, 1993: 801) and it is now universally accepted that Meinhof's so-called "Hamitic" group, which was based in part on the presence or absence of gender in the various languages, is in no way a valid linguistic taxon.

Recently Nichols has sought to resurrect the typological approach to historical linguistics in a book (1992) and series of articles (Nichols, 1990, 1995; Nichols and Peterson, 1996, 1998). According to Nichols (1992: 36), "one of the advantages claimed here for population typology is its ability to draw historical inferences from areal populations whose genetic classification is inadequate or incomplete from the perspective of standard historical work." The results of Nichols' enterprise, however, are no more valid than those of Meinhof. Space does not permit a complete examination of the numerous flaws in Nichols' work (Greenberg, 1993; Ruhlen, 1994a). I would, however, like to point out a few of the fatal flaws in what Nichols calls "population typology" and to explain why it cannot possibly lead to any "historical inferences."



inclusive/exclusive opposition in Australia, Oceania, Africa, and the Americas) constitute a global cline? For unrelated historical events to organize themselves into a global cline would seem to require either supernatural intervention or extremely good luck. In point of fact, however, the inclusive/exclusive opposition does *not* form a global cline. Rather, as Nichols herself admits, the inclusive/exclusive opposition is highest in Oceania, next highest in the Americas, and lowest in the Old World (Nichols, 1992: 206–7). This pattern thus in no way constitutes the east-to-west cline posited by Nichols. But this geographical inconvenience does not deter Nichols in her quest for global clines, for, according to Nichols, “water miles” are twice as long as “land miles” and therefore the Pacific Ocean is to be thought of as in the Atlantic Ocean and thus the pattern Oceania–Americas–Old World constitutes a global cline.

And “water miles” are not Nichols’ only discovery. She also attributes the presence of noun class systems in Africa not to the historical process of diachronic typology outlined in Greenberg (1978), but rather to the fact that Africa is a “hotbed” for noun classes and therefore it is to be *expected* that African languages should have noun classes. For Nichols, an explanation is simply a neologism away.

## 6. Conclusion

It is a sad commentary on the fate of historical linguistics in the twentieth century that families such as Eurasiatic and Amerind, whose outlines were already perceived by scholars such as Trombetti and Sapir at the beginning of the last century, are still considered controversial, at best, or completely spurious, at worst. Both families are in fact obvious if the proper basis of genetic classification — taxonomy — is understood and applied without preconception, rather than the caricature of the comparative method that became fashionable in the twentieth century. Equally deplorable was the retrogression in linguistic taxonomy during the twentieth century, where valid families that were well established and accepted

at the start of this century — Altaic, Na-Dene, Hokan, Penutian — came to be considered dubious by many traditional historical linguists. According to Hock and Joseph (1996: 498, 502), “Greenberg’s methodology of mass comparison must be considered of dubious reliability. . . . There is no credible alternative to the cumbersome and time-consuming traditional method of comparative linguistics.” Unfortunately this cumbersome and time-consuming method made no contribution to linguistic taxonomy in the twentieth century and it should be clear by now that it never will. I have tried in this article to clarify the methodological basis of genetic relationship, a methodology that was worked out in the nineteenth century by the pioneers of comparative linguistics and then largely forgotten or ignored during the twentieth century when the true principles of linguistic taxonomy disappeared from what became known as historical linguistics.

The time is past for historical linguists to continue to pretend that language can only be used to study human prehistory at very shallow time depths and that Indo-European represents the temporal limits of the comparative method. There is no evidence that this is true and there is a great deal that it is not. The reconstruction of human prehistory using comparative linguistics is not just a responsibility of the linguistic community, it is also a debt that we owe to the ancillary human sciences.

## References

- Ammerman, Albert J., and L. L. Cavalli-Sforza. 1984. *The Neolithic transition and the genetics of populations in Europe*. Princeton: Princeton University Press.
- Bynon, Theodora. 1977. *Historical linguistics*. Cambridge: Cambridge University Press.
- Campbell, Lyle. 1986. *Comments*. *Current Anthropology* 27.488.
- \_\_\_\_\_. 1994a. Inside the American Indian language classification debate. *Mother Tongue* 23.41–55.



- . 1994b. Putting pronouns in proper perspective in proposals of remote relationships among Native American languages. *Survey of California and other Indian languages, Report 8*, ed. by Margaret Langdon, 1–20. Berkeley: Survey of California and Other Indian Languages.
- . 1997. Amerindian personal pronouns: A second opinion. *Language* 73.339–51.
- Cavalli-Sforza, L. Luca. 1996. *Gènes, peuples et langues*. Paris: Odile Jacob.
- Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza. 1994. *The history and geography of human genes*. Princeton: Princeton University Press.
- Delbrück, Bertold. 1880. *Einleitung in das Sprachstudium*. Leipzig.
- Dimmendaal, Gerrit J. 1993. Review of On language: Selected writings of Joseph H. Greenberg. *Language* 69.796–807.
- Dolgopolsky, Aron B. 1964. Gipoteza drevnejshego rodstva jazykovyx semej severnoj evrazii s verojatnostnoj točki zrenija. *Voprosy jazykoznanija* 2.53–63.
- Durie, Mark, and Malcolm Ross. 1996. Introduction. *The comparative method reviewed: Regularity and irregularity in language change*, ed. by Mark Durie and Malcolm Ross, 3–38. New York: Oxford University Press.
- Dybo, Anna, Oleg Mudrak, and Sergei A. Starostin. 2003. *An Altaic etymological dictionary*. Leiden: Brill.
- Fox, Anthony. 1995. *Linguistic reconstruction: An introduction to theory and method*. New York: Oxford University Press.
- Giacone, Antonio. 1949. *Os Tucanos e outras tribus do Rio Uaupés afluente do Negro-Amazonas: notas etnográficas e folclóricas*. São Paulo.
- Goddard, Ives, and Lyle Campbell. 1994. The history and classification of American Indian languages: What are the implications for the peopling of the Americas? *Method and theory for investigating the peopling of the Americas*, ed. by Robson Bonnichsen and D. Gentry Steele, 189–207. Corvallis, OR: Center for the Study of the First Americans.
- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of Language*, ed. by Joseph H. Greenberg, 73–113. Cambridge: M.I.T Press.
- . 1971. The Indo-Pacific Hypothesis. *Current Trends in Linguistics, vol. 8*, ed. by J. Donald Bowen, et al., 807–71. The Hague: Mouton.
- . 1978. How does a language acquire gender markers? *Universals of human language, Vol. 3*, ed. by Joseph H. Greenberg, 47–82. Stanford: Stanford University Press.
- . 1981. *Amerindian Notebooks, 21 volumes*, unpublished.
- . 1987. *Language in the Americas*. Stanford: Stanford University Press.
- . 1993. Review of Linguistic diversity in space and time, by Johanna Nichols. *Current Anthropology* 34.503–5.
- . 1995. The concept of proof in genetic linguistics. *Mother Tongue* 1.207–16.
- . 2000–02. *Indo-European and its closest relatives, 2 vols.* Stanford: Stanford University Press.
- Haas, Mary. 1966. Wiyot-Yurok-Algonkian and problems of comparative Algonkian. *International Journal of American Linguistics* 32.101–7.
- Hock, Hans Henrich, and Brian D. Joseph. 1996. *Language history, language change, and language relationship: An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.
- Illich-Svitych, Vladislav M. 1971–84. *Opyt sraunenija nostraticheskix jazykov, 3 vols.* Moscow: Nauka.
- Kaufman, Terrence. 1990. Language history in South America: What we know and how to know more. *Amazonian linguistics: Studies in lowland South American languages*, ed. by Doris L. Payne, 13–73. Austin: University of Texas Press.
- Klein, Richard. 1999. *The human career*. Chicago: University of Chicago Press.
- Meillet, Antoine. 1965 [1914]. Le problème de la parenté des langues. *Linguistique historique et linguistique générale*, by Antoine Meillet, 76–101. Paris: Champion.
- Miller, Roy Andrew. 1971. *Japanese and the other Altaic languages*. Chicago: University of Chicago Press.
- . 1991a. Genetic connections among Altaic languages. *Sprung from Some Common Source*, ed. by Sydney M. Lamb and E. Douglas Mitchell. Stanford: Stanford University Press, 293–327.
- . 1991b. Anti-Altaicists contra Altaicists. *Ural-Altaische Jahrbücher* 63.5–62.
- . 1991c. How many Verner's Laws does an Altaicist need? *Studies in the Historical Phonology of Asian Languages*, ed. by William G. Boltz and Michael C. Shapiro, 176–204. Amsterdam: John Benjamins.
- Nichols, Johanna. 1990. Linguistic diversity and the first settlement of the New World. *Language* 66.475–521.
- . 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- . 1995. The spread of language around the Pacific rim. *Evolutionary Anthropology* 3.206–15.
- . 1996. The comparative method as heuristic. *The comparative method reviewed: Regularity and irregularity in language change*, ed. by Mark Durie and Malcolm Ross, 39–71. New York: Oxford University Press.



- Nichols, Johanna, and David A. Peterson. 1996. The Amerind personal pronouns. *Language* 72.336–71.
- \_\_\_\_\_. 1998. A reply to Campbell. *Language* 74.605–14.
- Proulx, Paul. 1985. Proto-Algic II: verbs. *International Journal of American Linguistics* 51.59–93.
- Rankin, Robert L. 1992. Review of *Language in the Americas*, by Joseph H. Greenberg. *International Journal of American Linguistics* 58.324–49.
- Ruhlen, Merritt. 1994a. Review of *Linguistic diversity in space and time*, by Johanna Nichols. *Anthropos* 89.640–41.
- \_\_\_\_\_. 1994b. First- and second-person pronouns in the world's languages. *On the origin of languages: Studies in linguistic taxonomy*, 252–60. Stanford: Stanford University Press.
- \_\_\_\_\_. 1994c. Amerind T'A?NA 'child, sibling.' On the origin of languages: *Studies in linguistic taxonomy*, 183–206. Stanford: Stanford University Press.
- \_\_\_\_\_. 1995a. A note on Amerind pronouns. *Mother Tongue* 25.60–61.
- \_\_\_\_\_. 1995b. Proto-Amerind numerals. *Anthropological Science* (Tokyo) 103.209–25.
- \_\_\_\_\_. 2002. *An index to Amerind reconstructions*, unpublished.
- Sapir, Edward. 1913. Wiyot and Yurok, Algonkian languages of California. *American Anthropologist* 15: 617–46.
- \_\_\_\_\_. 1923. The Algonkian affinity of Yurok and Wiyot kinship terms. *Journal de la Société des Américanistes de Paris* 15.36–74.
- Sarich, Vincent M. 1994. Occam's razor and historical linguistics. *In honor of William S-Y. Wang: Interdisciplinary studies on language and language change*, ed. by Matthew Y. Chen and Ovid J. L. Tzeng, 409–30. Taipei: Pyramid Press.
- Starostin, Sergei A. 1991. *Altajskaja problema i proisxozhdenie japonskogo jazyka*. Moscow: Nauka.
- Trombetti, Alfredo. 1905. *L'unità d'origine del linguaggio*. Bologna: Luigi Beltrami.
- Watkins, Calvert. 1990. Etymologies, equations, and comparanda: Types and values, and criteria for judgement. *Linguistic Change and Reconstruction Methodology*, ed. by Philip Baldi, 289–303. Berlin: Mouton de Gruyter.