



Classification of American Indian Languages: A Reply to Campbell

Joseph H. Greenberg

Language, Vol. 65, No. 1 (Mar., 1989), 107-114.

Stable URL:

<http://links.jstor.org/sici?sici=0097-8507%28198903%2965%3A1%3C107%3ACOAILA%3E2.0.CO%3B2-U>

Language is currently published by Linguistic Society of America.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/lsa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

DISCUSSION NOTE

Classification of American Indian Languages: A Reply to Campbell

JOSEPH H. GREENBERG

Stanford University

In *Language* 64:3 (1988), Lyle Campbell contributed a review article concerning my book *Language in the Americas* (henceforth C and *LIA*, respectively) that was uniformly negative. I believe that readers of *Language* should be made aware that one year before my book appeared in print, on the basis of a brief exposition of its contents in *Current Anthropology* (Greenberg, Turner, & Zegura 1986), Campbell wrote that my classification of American Indian languages 'should be shouted down' (1986:488). Under these circumstances, an objective review could hardly be expected. More importantly, the reader of C's review article gets no real notion of the main arguments and contents of the book.

I will first discuss several specific points made by C and then expand on two more general topics, those sections of my book (and the basic results contained in them) that C does not mention and the question of so-called Pan-Americanisms. C claims that my classification is not 'new' (591), referring to my 1956 paper (Greenberg 1960). But this paper was only four pages long, gave details only for South America and South American outliers in Central America, and presented no linguistic evidence. Even so, if it was not subsequently investigated by others, especially in regard to the new proposals for South America, this was not my fault, but stemmed from a combination of attitudes, essentially a relative lack of interest in the basic problem and a generally negative attitude towards all attempts at broad classification.

C denies that my success in African classification proves anything about the validity of my Amerindian work. This is, of course, true in the sense that it must stand on its own. Still, there should be some presumption that methods successful in one area will also be successful when applied elsewhere. These methods are set forth in great detail in Ch. 1 of *LIA*, but C does not even mention the existence of such a chapter. With regard to African classification, he says that parts are in dispute, but does not specify which. In fact, my classification is clearly the basis of present-day African historical linguistics. There is no alternative classification, and the few proposals cutting across my four basic groups have received no general support. All disputes have been at the level of subgroupings, some of which I had said were tentative; and indeed some of the changes were suggested by me. When one considers that even today there is no unanimity regarding Balto-Slavic as a subgroup of Indo-European (IE), similar disputes among the much less studied African languages should come as no surprise. A Nilo-Saharan newsletter and a *Journal of Afroasiatic Linguistics* exist, and a comparative dictionary of Afroasiatic, with no alteration of membership in the family and only one important subgrouping

change, is proceeding under the editorship of Diakonov (1981–). There is also my Australian classification (*LIA*, 29), with which one may compare Dixon (1980:20)—which appeared later and independently.

C asserts that ‘the genetic relationship among putative members of Na-Dene itself has been seriously questioned’ (593). He cites Levine 1979, but fails to mention that an entire chapter of *LIA* was devoted to a refutation of Levine’s critique of Sapir’s inclusion of Haida in the Na-Dene family. In the 1987 *Current Anthropology* review of *LIA*, two respondents (Hymes, 662, and J. D. Sapir, 664) called this refutation ‘brilliant’. Campbell & Mithun 1979 endorsed Levine’s article emphatically, and Michael Krauss, in the same volume, said that we should be grateful to Levine for debunking Sapir’s thesis once and for all.

Sapir presented two-way etymologies between Haida and Athabaskan, and three-way etymologies among Haida, Tlingit, and Athabaskan. Levine cites seven criteria, most of them in my view irrelevant, for rejecting these etymologies. Levine repeatedly rejects the same etymology by citing different criteria. It apparently occurred to no one, in their eagerness to show that Sapir was wrong, even to note how many etymologies survived all attacks: 17 of the three-way etymologies and 14 of the Haida-Athabaskan comparisons. In my book I compared three branches of IE, applying Levine’s criteria to the etymologies involving these three branches in Pokorny’s standard comparative dictionary. Only six three-way comparisons survived when I used some latitude with regard to IE vowel alternation. If the criteria were strictly applied, none of the etymologies would survive.

C calls the proposed affiliation of Xinka and Lenca a ‘long-shot proposal’ (594), and he cites and attacks the evidence in Lehmann (1920:767). I noted Lehmann’s proposal, just as I did many others; some I accepted, some I did not. But what Lehmann and many others have done—comparing just two languages—is exactly what I never do, since it does not lead to an overall classification. To me Xinka and Lenca are related because they are both Chibchan. Any evidence showing that either is Chibchan is relevant. Since Chibchan is a large group, there are innumerable pairs one could compare in isolation, and it would make no more sense than if an Indo-Europeanist first tried to show that Lithuanian was related to Rumanian, then Greek to Albanian, etc. He would be excluding much relevant evidence, and even if he proved all of them he would still not distinguish the IE family. In fact, Xinka and Lenca have no special relationship within Chibchan. Xinka forms occur in 33 Chibchan etymologies and 12 Amerind etymologies. The figures for Lenca are similar. In some instances both languages are found; of these, some were cited by Lehmann and some I discovered myself.

C claims that ‘most *LIA* examples are lexical’ (596). He generally ignores an entire chapter devoted to grammatical evidence. In fact, the line between grammatical and lexical evidence is not always clear because of the process of grammaticalization. Furthermore, I have stressed the importance of agreement in morphophonemic alternations as being almost surely genetic in origin, particularly in cases of arbitrary suppletion. In *LIA* I gave an example of this (46–49), namely, the occurrence in widely separated areas of third person *i-*

before consonants alternating with *t*- before vowels. The obscure notions about 'deep grammar' found in the American Indian linguistic literature are of three basic types: (1) agreement in alternations, particularly arbitrary suppletions; (2) agreement in whole or large parts of paradigms (e.g. Sapir's pronominal evidence for Algic); and (3) the marginal survivals in only a few words of grammatical elements that are productive elsewhere. As an example of this latter category one may cite the survival of the Chibchan-Paezan general numeral classifier *kwa*- (originally used for small round objects) on a few numerals in some Chibchan languages (*LIA*, 298–99). Like Sapir, and like the pioneers of IE comparative grammar, I use both grammatical and lexical evidence. The preceding example (and many others that could be cited from my book) are surely examples of 'deep grammar' in any reasonable sense of the word.

C's article contains a serious misunderstanding of my views regarding glottochronology. I agree with C that it is a misuse of glottochronology to employ it as a device to classify languages, and I have never used it that way. In fact I have pointed out additional weaknesses of glottochronology, in *LIA* and other publications, both for classification and subgrouping. The point of Appendix A, discussed in *LIA* (28–29), is quite different. It supposes that by other methods we already have distinguished a valid linguistic stock. Then, based on Brugmann's rule that an etymon is recoverable for the proto-language if it occurs in two or more separate branches, it is clear that the more ramified a language stock is, and the shorter the period of separation of the branches, the more of the original vocabulary is recoverable. The whole stock can then enter into still deeper comparisons, utilizing quite a large portion of its original vocabulary. The whole argument is directed against 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.

C suggests that I may have learned about mass comparison from Alfred Kroeber (597). This guess happens not to be true. In my discussions with Kroeber, who taught at Columbia after his retirement and whose colleague I was, we discovered that we had been using the same method. As for my late discovery of mass comparison, in my first classification of African languages I proposed 16 stocks. This was mainly because linguists had not yet seriously investigated the Eastern Sudan in detail. In this first classification I distinguished stocks about on the level of Algic. I noted many resemblances among them in ways that crisscrossed and led to no clear results. Finally it struck me one day that nothing essential changed methodologically at higher levels. Hence I proceeded to compare all of the African stocks, just as I had previously compared individual languages. I found that most of those in the Eastern Sudan belonged together in a family I named Nilo-Saharan, a grouping now universally accepted.

C criticizes McQuown's 1942 Macro-Mayan hypothesis (598) because one of McQuown's reasons for positing it, the existence of glottalized consonants in Totonac, holds for other languages of the area outside of Macro-Mayan. I accepted Mexican Penutian, of which Macro-Mayan is a major part, for my

own reasons. As in the case of Lehmann, C is criticizing another linguist, whose work I certainly appreciated, but whose views I adopted, at least partially, for reasons of my own. It is obvious to anyone who knows my work that I would immediately rule out glottalization as an irrelevant typological feature.

C suggests (597) that I put together my notebooks in accordance with a predetermined classification that was reflected in the notebooks as assembled. But where did I get the 'predetermined classification'? My procedures are clearly described in Greenberg 1960, to which C refers. I started out by comparing about 40 words in a large number of South American languages and finding that they fell into several clearly marked groups, such as Andean and Macro-Ge. Only then did I set up notebooks. When I found new languages they either belonged to one of the previously recognized groupings, whose characteristics became clearer the more languages I used, or formed new groups by themselves.

C accuses me of failing to detect borrowings (599). In many instances I did detect them, in which case I simply omitted the form in question. In other instances I point to them as possible. For example, in the Amerind etymology for 'liver' (240) I noted that some Chibchan-Paezan forms might be borrowed from Spanish *pecho*, and in grammatical section 34 of Ch. 5 I suggested that the pronominal plural *-to* (found only in Fulnio among Macro-Ge languages) might be borrowed from a Macro-Panoan language, a group in which it is well attested. Doubtless I have made some errors in eliminating borrowings, but such inevitable flaws in a pioneering work such as *LIA* can hardly be taken to invalidate the book's basic thesis, the unity of the Amerind family.

In his discussion of the semantics of the etymologies, C finds that 'G's forms are quite permissive in semantic latitude' (600). I believe that I have been extremely sober. As a test, consider the glosses for the Almosan-Keresiouan etymologies 10, 20, etc., up to 100, with the numerals in parentheses indicating the number of instances of a particular gloss in the same etymology: 10: ask (3); 20: belly (4), breast (1); 30: break (1), be hit (1); 40: buy (1), trade (1), buy, sell (1), take (2); 50: come (1), arrive (2) go (1); 60: egg (2), bird's egg (1); testicle (1); 70: finish (4), finish, make, do (1); 80: foggy (1), moist (1), smoke (1), 90: grandfather (6), old man (1); 100: hit, fall into, slap (1), hit (1). I believe that this compares favorably with just about any etymological dictionary. C has taken a few complex etymologies and omitted the connecting semantic links. For instance, C's first example of permissive semantics, 'excrement/night/grass', is actually taken from the Amerind etymology for 'dirty' (*LIA*, 212). When examined in detail this etymology shows great semantic coherence; not only is it a strong etymology in its own right, but it even shows good evidence for the subgroups of Amerind proposed in *LIA*. A better gloss would have been 'black', or perhaps 'dark in color' for the original meaning. It occurs in five subgroups of Amerind. In Almosan its meaning is uniformly 'black', and the distributions throughout North and South America suggests that this was probably the original meaning. Keresiouan shows both 'dark in color' (Iroquoian) and 'green' (Keresan). In Penutian the meaning has shifted completely to 'green' and its close semantic relatives 'grass' and 'blue'. In South America,

Macro-Tucanoan shows the meaning 'black' everywhere, except for Canichana 'night' and Shukuru 'Negro'. In Macro-Ge, the original meaning is preserved in Proto-Ge 'black', but in Cayapo and Chiquito the meaning has shifted to 'dirty'. Finally, in the Equatorial group the meaning is uniformly 'excrement'. Remarkable evidence for the validity of the Penutian grouping is found in the shared semantic shift to 'green' and the reduplicated forms in three main subgroups, Plateau, California, and Mexican (North Sahaptin *tʰəktʰək*, Rumsien *čuktuk*, and Zoque *tʰuhtʰuh*). C has simply omitted all the connecting links. In Pokorny's standard comparative IE etymological dictionary it is easy to find more drastic examples; for instance, under the root *perk-* one finds the meanings 'ask', 'temple', 'prophecy', 'herald'.

Some of my etymologies C rejects as onomatopoeic. Here as elsewhere C does not realize that an etymological dictionary is not meant as a 'proof' of relationship. Some items are of course more cogent in this respect than others. However, all languages have onomatopoeic expressions, including proto-languages. Hence there are in Pokorny, as in every etymological dictionary I have ever seen, some onomatopoeic expressions.

In his discussion of submerged grammatical peculiarities—which are never clearly defined—C points out that even these may be treacherous; he gives two examples (601). One is an apparently arbitrary phonological alternation in which Proto-Mayan resembles Quechua. The unwary reader might think that I had proposed it. It involves, along with other forms, Quechua *-ni* as first person singular, stated to be merely a connecting morph. In fact, in my lengthy listing of *n-* markers for the first person (usually singular) in Amerind languages, I myself eliminated this item, which does not figure in my discussion (*LIA*, 49–50) precisely because internal Quechua data show that it is not a first person marker. C's second example concerns a discontinuous negative in Quechua A and Quiché, a Mayan language. Again I did not suggest this. C notes that the second component in Quiché, *ta*, is historically secondary. But one can tell this by bringing Mayan as a whole into the comparison, which is exactly the method I use. I not only did not suggest this (and C does not claim I did), but more importantly the methods I use would eliminate, and in fact did eliminate, both without my mentioning them.

C uses one device, a comparison of Amerind with Finnish, with regard to both my grammatical and lexical etymologies. His point is to show that my results are random by taking Finnish and trying to show that if I had encountered it in South America I would have classified it as an Amerind language. But 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. In *LIA* (24) there is a table of nine basic words from 25 European languages. The fact that Finnish is closest to Estonian, and that their closest relative is Hungarian, and that this group is distinct from IE and Basque appears, literally, from the first word on. It is a group at this level that should be compared with Amerind, and once more its distinctness is obvious. The large majority of C's forms do not even make it to Hungarian.

C claims that 'the number of the expected accidental matchings will be

roughly proportional to the number of languages consulted' (603). It is true that if one adds another language at random there will be an increase in the number of resemblances. However, C fails to ask how many will be found in more than two languages. It is an elementary proposition of probability theory that the probability of multiple accidents is a product of their individual probabilities. Since, of course, we are dealing with fractions, this involves a decrease. The likelihood of finding a resemblance in sound and meaning in three languages is the square of its probability in two languages. In general the probability must be raised to the $n - 1$ power for n languages. Thus, if five languages each showed a total of 8 percent sound-meaning resemblances pairwise, one would expect on a chance basis approximately 1/25,000 in all five languages. This multiplication of probabilities is discussed in Greenberg (1957:39), but was earlier noted by Collinder 1949 and no doubt by others as well. More concretely, if I have a group like the Western Romance languages (Italian, French, Spanish, Portuguese), there is an enormous difference between adding Rumanian and adding Basque. If I add Rumanian many three-way resemblances become four-way, etc., and a fair number of new etymologies appear. If I add Basque almost nothing happens.

The remainder of this reply addresses alleged inaccuracies of citation and analysis in *LIA*. C says that nearly all Americanists find shocking distortions in my data. However, Hymes (1987) noted, in regard to Ch. 2 of *LIA*, that '... from the standpoint of Chinookan it is sometimes clearly confirming, sometimes capable of addition, once or twice questionable.' In spite of this and other comments of a like nature, there are of course errors. As I stated in *LIA* (ix), 'the present work is in many respects a pioneering one... 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. I will be grateful for any corrections suggested by readers. However, I believe the work should be judged as a whole ... any user of dictionaries, even of intensively studied language families, will encounter numerous instances in which the same form has been assigned to different etymological entries by different scholars, or even in which the same form has erroneously been included in different etymologies.'

Little has been said in this reply regarding regular sound changes and correspondences. The question was discussed at considerable length in Ch. 1, which C hardly mentions in his review. What happens when one merely asks whether there is a relationship and then applies the rigid and largely irrelevant criteria used by Levine with regard to Haida and Na-Dene has been shown above. If applied to IE it would quite simply destroy IE comparative linguistics, which is in practice the main branch of historical linguistics. C apparently approves of Levine's methods and results, and similar statements of method by Kaufman and Goddard are even more restrictive. To see how unrealistic they are, the reader may peruse the list of French and English cognates (derived from etymological dictionaries) given in *LIA* (20-21). Using such methods one could disprove the affiliation of Hittite with Indo-European, whereas, in fact, Hittite revolutionized our ideas about the Proto-Indo-European sound system.

From all of C's discussion it would appear that the data assembled by me

are so defective and misleading that we must conclude that there are no significant resemblances among the 200-odd independent stocks he posits for the Americas. But this runs against the doctrine of Pan-Americanisms, the first reference to which I find in Campbell & Kaufman (1981:853), where they talk of 'widespread forms (so-called Pan-Americanisms)'. They use this concept in their criticism of Brown & Witkowski's 1979 article on Mayan-Zoque, a treatment restricted to proto-velars. Campbell & Kaufman assert that any attempt to prove that Mixe-Zoque is related to Mayan must exclude 14 of the proposed etymologies because they are Pan-Americanisms. If Pan-Americanisms are to be considered genetically related forms, this is of course contrary to normal practice. In an etymological dictionary of Germanic, no one excludes forms like English *two* and German *zwei* because this is an IE etymon.

There is no mention in C of the widespread forms *n-* 'first person' and *m-* 'second person'. *LIA* (49–55) contains what I believe is the first detailed enumeration of the distribution of these pronouns, which extend from British Columbia to Chile and occur in every subgroup of Amerind. This distribution cannot be explained either by borrowing or chance. The borrowing of first- and second-person pronouns is very rare. That a highly improbable event should have recurred more than a hundred times exceeds the bounds of credibility. A number of widespread grammatical patterns discussed in Ch. 5, which include irregularities, also cannot be explained plausibly except as the result of genetic inheritance. One wonders why scientists, who should be impartial regarding types of explanation, should avoid genetic explanations in such cases in favor of borrowing over a distance far greater than that covered by IE, and which would require contacts of virtually every language with every other one, given the variety of distributions.

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 (Greenberg, Turner, & Zegura 1986). My linguistic classification was arrived at in total independence of this external evidence, and until recently I was unaware of the agreement. Of course it is not probative, but it is still remarkable and should be of interest to readers of *Language*. A recently completed world-wide study of a large number of genes by Cavalli-Sforza and his associates (1988) not only confirms Zegura's results, but shows that from a world-wide perspective the biological differences between North and South American Indians are minimal, if we exclude Na-Dene and Eskimo-Aleut, which I classify separately.

REFERENCES

- BROWN, CECIL H., and STANLEY R. WITKOWSKI. 1979. Aspects of the phonological history of Mayan-Zoquean. *IJAL* 45.34–47.
- CAMPBELL, LYLE. 1986. Comment on Greenberg, Turner, and Zegura. *Current Anthropology* 27.488.
- . 1988. Review article on *Language in the Americas*. *Lg.* 64.591–615.
- , and TERRENCE KAUFMAN. 1981. On Mesoamerican linguistics. *American Anthropologist* 82.850–7.

- , and MARIANNE MITHUN (eds.) 1979. *The languages of Native America: Historical and comparative assessment*. Austin: University of Texas Press.
- CAVALLI-SFORZA, L. L.; ALBERTO PIAZZA; PAOLO MENOZZI; and JOANNA MOUNTAIN. 1988. Reconstruction of human evolution: Bringing together genetic, archeological and linguistic data. *Proceedings of the National Academy of Sciences* 85.6002–06.
- COLLINDER, BJÖRN. 1949. *La parenté linguistique et le calcul des probabilités*. Uppsala universitets årsskrift 13.
- DIAKONOV, I. M. (ed.) 1981– . *Sravnitel'no-istoričeskij slovar' afrazijskix jazykov*. MOSCOW.
- DIXON, R. M. W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.
- GREENBERG, JOSEPH H. 1957. *Essays in linguistics*. Chicago: University of Chicago Press.
- . 1960. The general classification of Central and South American languages. *Men and cultures: Selected papers of the 5th International Congress of Anthropological and Ethnological Sciences, 1956*, ed. by Anthony Wallace, 791–94. Philadelphia: University of Pennsylvania Press.
- . 1987. *Language in the Americas*. Stanford: Stanford University Press.
- ; CHRISTY G. TURNER II; and STEPHEN L. ZEGURA. 1986. The settlement of the Americas: A comparison of the linguistic, dental and genetic evidence. *Current Anthropology* 27.477–97.
- HYMES, DELL. 1987. Review of *Language in the Americas*, by Joseph H. Greenberg. *Current Anthropology* 28.659–62.
- KRAUSS, MICHAEL. 1979. Na-Dene and Eskimo-Aleut. In Campbell & Mithun 1979, 803–901.
- LEHMANN, WALTER. 1920. *Zentral-Amerika*. Berlin: Museum für Völkerkunde zu Berlin.
- LEVINE, ROBERT D. 1979. Haida and Na-Dene: A new look at the evidence. *IJAL* 45.157–70.
- MCQUOWN, NORMAN. 1942. Una posible síntesis lingüística macro-mayence. *Mayas y olmecas, Segunda Mesa Redonda sobre problemas antropológicos de México y Centro América*, 37–8. Mexico: Sociedad Mexicana de Antropología.
- POKORNY, JULIUS. 1959. *Indogermanisches Etymologisches Wörterbuch*. Bern: Francke.
- SAPIR, J. DAVID. 1987. Review of *Language in the Americas*, by Joseph H. Greenberg. *Current Anthropology* 28.663–64.

Department of Anthropology
Stanford University
Stanford, CA 94305

[Received 8 November 1988;
revision received 16 November 1988;
accepted 17 November 1988.]