Binary Comparison and the History of Hokan Comparative Studies

William J. Poser Stanford University

[This paper appeared in 1995 in the *International Journal of American Linguistics* **61**.1.135-144. This is the final text as transmitted to the journal. Details of formatting and pagination differ.]

1. Introduction

In recent statements on the methodology of linguistic classification, Joseph Greenberg and Merritt Ruhlen have argued that it is essential to compare many languages simultaneously rather than in pairs, as they believe to be the common practice. Greenberg (1987;132) holds up the Hokan family, whose membership and subgrouping are problematic, as a prime example of the impediment to progress presented by "binary comparison", with which he contrasts his method of "multilateral comparison":

The present work has profited from several studies, notably Jacobsen (1958) on Washo and Karok; Silver (1964) on Shasta and Karok; McLendon (1964) on Eastern Pomo and Yana; and Oltrogge (1977) on Jicaque, Subtiaba, and Tequistlatec. But despite these and numerous other two- or three-way comparisons of Hokan languages published since Sapir's day, the etymologies presented below, and the Hokan entries in the Amerind dictionary, contain a considerable number of new items or new entries under already published ones. Here, as clearly as anywhere, do we see the limitations of starting from one particular language or comparing only a few languages. By such a procedure, even the most obvious etymology involving Pomo, Chumash, and Tequistlatec would not appear in any published study because it does not involve Yana (Sapir), Subtiaba (Sapir), Jicaque (Swadesh and Greenberg), and so on.

Greenberg's complaint is echoed by Ruhlen (1987;234):

¹ I am grateful to an anonymous referee for his or her detailed comments on the draft of this paper.

... no Amerindian group has suffered so grievously from binary comparison as Hokan. Since a binary approach to language classification is useless in determining subgrouping, it is not surprising that there has been no progress in unraveling the internal relationships of Hokan in the past half-century.

Greenberg and Ruhlen seem to be concerned both with membership in the family and with subgrouping. As far as subgrouping is concerned, there is surely no quarrel — using classical methods (shared innovations), subgrouping is impossible if only pairs of languages are studied since it is necessary to reconstruct the proto-language and work out the history of each daughter language. Since neither Greenberg nor Ruhlen has described what methods he would use for subgrouping languages, and since neither has offered any evidence regarding the subgrouping of Hokan, we have no way of evaluating the putative superiority of their methods. In any case, the subgrouping of Hokan presented by Greenberg (1987;380-381) and Ruhlen (1987;369) so closely resembles the subgroupings to be found in the literature (e.g. Sapir 1929) that one wonders what advantage Greenberg and Ruhlen might have in mind. In the remainder of this note I therefore put aside subgrouping and concentrate on genetic affiliation

Neither Greenberg nor Ruhlen expands on his view of the history of Hokan comparative studies, so it is not clear precisely where the damage they perceive has been done. What their statements lead us to suspect is that affiliation of the various Hokan languages has been postulated on the basis of comparisons of a small set of pairs of languages, with the family as a whole recognized only by transitivity (that is, language A compared with B and B with C, but never A directly with C), and that subsequent work has taken the same form, with studies restricted to particular pairs of languages.

If this is what is meant by the claim that Hokan has suffered from binary comparison, a review of the history of work on Hokan makes it very clear that Hokan has **not** suffered from binary comparison; in fact, there has been very little binary comparison.² Greenberg and Ruhlen's complaint mischaracterizes the Hokan comparative literature and misattributes the problematic status of Hokan.

2. Overall Surveys

However many binary studies there may have been, the existence of overall surveys of Hokan renders Greenberg and Ruhlen's point moot. Two of Sapir's

² This point emerges clearly from the review of Hokan classification in Jacobsen (1979). Kaufman (1988;57) appears to agree with Greenberg and Ruhlen when he says: "Essentially, apart from Sapir, Swadesh, Bright, Gursky, and Greenberg, scholars generally have not accepted Hokan-Coahuiltecan as a whole or at least have not tried to test the hypothesis. Rather, they have carried out binary studies involving one or more languages from each of two genetic groups, or they have limited themselves to the languages of California." Note, however, that he excepts Bright and Gursky, and that he recognizes that many scholars have engaged in non-binary work on languages of California, which comprise the majority of the languages of the putative Hokan family.

papers, Sapir (1917a) and Sapir (1925), provided overall surveys of Hokan. The most recent and comprehensive surveys of Hokan are Gursky (1974), Langdon (1974), Jacobsen (1979), and Kaufman (1988). Of these, Gursky presents the most extensive set of equations.³ Haas (1963) is a smaller scale study that discusses in detail nine equations with members drawn from all 13 branches of Hokan.

Gursky (1974) presents hundreds of equations, covering all of the putative Hokan languages, and while he is acquainted with the previous comparative literature, he includes many equations derived from his own comparisons. As a result, his work, and therefore Hokan, certainly cannot be said to suffer from the disadvantages of binary comparison. Indeed, Greenberg's claim that any equation involving Pomo, Chumash, and Tequistlatec will have been missed in previous work on Hokan is belied by Gursky's paper. Gursky (1974;207)'s equation STONE (2) involves the following languages: Proto-Pomo, Southern Pomo, Proto Southwest Pomo — Central Pomo, Barbareño Chumash, Tequistlatec (Coastal Dialect), and Tequistlatec (Highland Dialect), as well as Chimariko, Esselen, Kiliwa, and Proto-Yuman, thus counterexemplifying Greenberg's claim.⁵

In addition to such surveys of the entire Hokan family, there are papers that deal with putative subgroups. An example is Bright (1954), which discusses subsubgroup A of Sapir's Northern Hokan subgroup, comprising Karok, Chimariko, Shasta, Achumawi and Atsugewi.

3. How the Hokan Family Was Assembled

Greenberg and Ruhlen's complaint suggests that the Hokan family is the result of a number of binary pairings of languages which, if accepted, result by transitivity in the family. A review of the process by which the Hokan family was assembled belies this suggestion. The following is a concise summary of the construction of the Hokan family.⁶

Brinton (1891)

Argued for the affiliation of Yuman, Seri, and Tequistlatec.

Dixon (1905)

³ Kaufman (1988) takes a more conservative position than Gursky (1974) on the membership of the Hokan family, and attempts to set up phonological correspondences, rather than merely listing resemblant forms as Gursky and Greenberg do. It thus represents the state of the art in Hokan comparison and reconstruction. However, since it appeared after Greenberg (1987) and Ruhlen (1987), I will not discuss it further here.

⁴ Gursky (1988,1989,1990) provides additions and corrections to Gursky (1974).

 $^{^5}$ As it happens, Greenberg's equations for STONE do not contain this set as he splits the forms cited by Gursky into two different equations.

⁶ For more detailed history see Langdon (1974) and the references cited by Jacobsen (1986;46, fn.1).

Grouped Shasta and Palaihnihan (Achomawi and Atsugewi).

Dixon (1910)

Related Chimariko to Shasta and Palaihnihan.

Dixon & Kroeber (1913a)

Added Karok, Pomo, Esselen and Yuman to Chimariko, Shasta, and Palaihnihan and named the resulting family "Hokan".

Dixon & Kroeber (1913b)

Proposed the Iskoman group, consisting of Salinan and Chumash, and suggested the addition of Iskoman to Hokan.⁷

Kroeber (1915)

Added Seri and Tequistlatec.

Swanton (1915)

Created the Coahuiltecan family, consisting of: Coahuilteco, Cotoname, Comecrudo, Karankawa, Tonkawa, and Atakapa.

Sapir (1917a)

Added Yana.

Sapir (1917b)

Joined Coahuiltecan and Hokan, to form Hokan-Coahuiltecan, often referred to as Macro-Hokan or simply as Hokan. His comparisons included the following 17 languages and families: Atakapa, Chimariko, Chumash, Coahuilteco, Comecrudo, Cotoname, Esselen, Karankawa, Karok, Pomo, Salinan, Seri, Shasta, Tequistlatec, Tonkawa, Yana, Yuman

Dixon & Kroeber (1919)

Added Washo.

Lehmann (1920)

Added Subtiaba.

Sapir (1921)

Presented additional equations of Salinan with Hokan and Coahuiltecan.

Sapir (1925)

Advocated the addition of Subtiaba, previously proposed by Lehmann (1920).

Rivet (1942)

Gatschet (1877) regarded the Antoniaño dialect of Salinan as a Chumash dialect, thereby implicitly proposing Iskoman, but gave no evidence or argument.

Added Yurumanguí.

Greenberg & Swadesh (1953)

Added Jicaque.

It is easily seen from the above summary that the Hokan family was assembled by a process of accretion of languages to the already established core and of combination of whole groups of languages into still larger groupings, not by a set of binary comparisons.

4. Putative Binary Comparisons

Recent work on Hokan has in a number of cases taken the form of papers that, at first glance, involve binary comparison. The binary comparisons to which Greenberg and Ruhlen presumably refer are the following:

Palaihnihan and Shasta Olmsted (1956,1957, 1959) Jacobsen (1958) Washo and Karok Haas (1964) Yana and Karok McLendon (1964) Eastern Pomo and Yana Silver (1964) Shasta and Karok Turner (1967) Seri and Tequistlatec J. M. Crawford (1976) Chimariko and Yuman Seri and Yuman J. G. Crawford (1976)

These papers were all motivated by the availability of improved data. For example, Olmsted's papers were motivated by his own field work on Palaihnihan, and Jacobsen (1958) was motivated by the new data made available by his own fieldwork on Washo and William Bright's field work on Karok.

These binary comparisons have **not** led to the conclusion that the languages are not related. In fact, every one of these studies but Turner (1967), concludes that the languages considered are related. Indeed, in the case of Olmsted's papers, the genetic affiliation of Palaihnihan and Shasta was never in question. His purpose was to improve the comparative phonology of the languages studied and to test the hypothesis that Atsugewi, Achumawi, and Shasta are more closely related to each other than to other northern Hokan languages. Thus, if the defect of binary comparison is that it may lead to an inappropriate failure to detect relationship, it cannot be said that this problem has manifested itself in the case of Hokan.

Moreover, even papers devoted primarily to the comparison of two languages have done their best to give comparisons with the entire family, a point previously noted by Voegelin & Voegelin (1977;158) and Jacobsen (1979;557). For example, in addition to giving numerous comparisons between Washo and Karok, Jacobsen (1958) gives additional comparisons to other Hokan languages, not only adding Washo and Karok forms to existing equations, but in some cases proposing new

comparisons between forms from other languages. Here is his own description of his work (Jacobsen 1958;195):

Besides a number of new two-way comparisons, it has been possible to find several comparisons between words in at least three branches of the family, no two of which had previously been compared.

He continues (Jacobsen 1958;204):

Some of the references are meant to suggest that the Washo and Karok words compared may be related to other words not previously compared to either, so that these together would form a more widely attested family of cognates ... Others will serve to add a Washo or Karok form to a set of cognates that is already well represented ... On the other hand, some references will show previously published juxtapositions equivalent to mine ... Yet others will show that the two words compared are part of a large and accepted set of cognates.

Similarly, J. G. Crawford (1976), which as its title suggests is devoted especially to the comparison of Seri with the Yuman languages, gives comparisons with other Hokan languages throughout the list of 227 equations.

Olmsted (1956, 1957, 1959), though devoted primarily to Atsugewi, Achumawi, and Shasta, brings in data not only from Karok, Chimariko, and Yana, other Northern Hokan languages, but also from more distantly related languages such as Salinan and Esselen.

Haas (1964;76) also included comparisons with languages other than Yana and Karok: "Problems of particular interest are further amplified by the addition of other Hokan cognates or probable cognates."

Yet another example is Waterhouse (1976). Devoted primarily to comparison of the two dialects of Chontal and reconstruction of Proto-Chontal, it also gives comparisons of Chontal with a wide range of languages, including not only recognized Hokan languages, such as the Yuman languages, Seri, and Karankawa, but also more controversial languages, such as Quinigua, Waikuri, and Tlappanec. She adds Chontal data to equations proposed by Gursky (1968) and Jacobsen (1958). Jacobsen (1979;567-70), inspired by Waterhouse (1976), offers Washo corrections and additions to the equations in Gursky (1974).

Indeed, attention to the family as a whole has meant that even comparisons of a single language with the core languages have not ignored equations not involving the language of interest. Sapir (1917a) not only added Yana forms to existing equations, but also gave (pp.26-27) six new equations not involving Yana. Greenberg & Swadesh (1953) included 15 equations that did not involve Jicaque.

5. Conclusion

The history of Hokan comparative studies reveals no evidence that research has suffered from binary comparison as alleged by Greenberg and Ruhlen. Instead, what we find is a gradual accretion of new languages to the previously proposed core, with comparison of the new language or languages with all of the core languages. Instances of binary comparison are few, sometimes involve putative subgroups, have typically included comparisons beyond the two languages on which they focus, and have in any case not led to rejection of genetic affiliation.

I submit that the Hokan family has been problematic for reasons quite unrelated to "binary comparison", namely the very limited data available for many languages and the nature of the evidence offered for genetic affiliation. Many of the links were originally posited on the basis of very slim evidence, consisting of unsystematic similarities in a very small number of words.

Consider, for example, the basis on which Salinan and Chumashan have been included in Hokan as members of the Iskoman subgroup. Here in its entirety is the evidence given by Dixon & Kroeber (1913b;652-653) for the Iskoman family and its inclusion in Hokan:

An apparent structural similarity of Chumash and Salinan was long ago noted by the authors, but as in the case of Yurok and Wiyot, lexical resemblances, while occurring, are to date not conspicuous. A presumption favorable to relationship may however be properly entertained on the basis of existing knowledge.

	Chumash	Salinan
water	o, to	t- a , t š- a (ocean)
rabbit	qun	kol (jack-rabbit)
jack-rabbit	ma	map (rabbit)
arm	pu	-ipokou
sky	alapa	$lem,\ lemak$
coyote	$alax\ddot{u}wul$	elka
stone	$x\ddot{o}p$	$-xap$, $t\check{s}$ - xa
\log	$hut\check{s}u,\ wut\check{s}u$	$ot \check{s}o$
ground squirrel	emet	- $emko$ ' m
two, four	$i\check{s}kom,\ \check{s}kumu,\ paksi$	$ki\check{s}a,\;kaki\check{s}e$
ten	tuyimili	tsoe
sixteen	$peusi, \ peta$	$kpe\check{s}$

Several of the above words lend themselves to the hypothesis of a connection between Hokan and Iskoman: water, house, sky, stone, two. To these may be added blood, Chumash axulis, Hokan ax-; no, Chumash pwo, Hokan po-; tongue, Salinan paL, Hokan p-l; salt, Salinan akai, Hokan aki, asi.

It is however idle to discuss further a possible relationship between Iskoman and Hokan, when the genetic connection between the members of Iskoman is scarcely yet a matter of demonstrable proof, probable though it may seem.

The relationship between Salinan and Chumashan is based on a total of twelve equations, few of them particularly striking, without either phonological correspondences or any account of the differences between putatively related forms, and no morphological evidence at all. The evidence for the connection with Hokan is slimmer still.

No further evidence has been adduced for a particularly close connection between Chumashan and Salinan. That Iskoman persists in spite of the extreme paucity of evidence in favor of it and the proposal of Haas (1963;57) that it be abandoned shows how little basis there is for parts of the Hokan classification.

Nor has much evidence been adduced for including Chumashan in Hokan. Indeed, a number of scholars, including Kaufman (1988) and the present author, believe that Chumashan is not properly included in Hokan. It is because far too much of Hokan rests on such meagre evidence that the family has been so problematic.

Many of the putative Hokan languages are poorly documented, or have been until relatively recently. As more fieldwork has been done, and as the extensive materials left by John Peabody Harrington on languages of California have been exploited, more and better data has become available for some languages, such as Karok, Washo, Salinan, and the Chumashan languages. Other poorly documented languages can be expected to remain so due to their early extinction. These include Esselen, the "Coahuiltecan" languages, the languages of Baja California, and Yurumanguí.

Even where good data is available, it has proved difficult, in many cases, to establish regular phonological correspondances based on significant sets of cognates, and relatively little morphological evidence has been found. Where such systematic correspondences have not been established, possible borrowings have not been carefully distinguished from inherited forms, as pointed out by Jacobsen (1979) with regard to Gursky (1974). The reason that the Hokan family has been and remains controversial is that for many of the links that make up the overall family, evidence of genetic affiliation of the sort considered probative by most historical linguists has been unacceptably thin. It is the tenuousness of the relationship of the putative Hokan languages and the lack of properly worked out derivations from Proto-Hokan that have made it impossible to detect the shared innovations that form the basis for classical (non-lexicostatistical) subgrouping.

Ironically, the real problems with Hokan classification are due to the very causes that make Greenberg's methodology questionable. What is controversial about Greenberg's methodology is not his comparison of many languages. It is his failure to establish the systematic correspondences between languages that remove the possibility of chance similarity, his lack of concern for loans, and his willingness to

postulate relationship on the basis of minute amounts of evidence,⁸ precisely the weaknesses that affect parts of Hokan.

Greenberg and Ruhlen's view that the Hokan family has suffered from binary comparison has no basis in fact. The problems in Hokan classification are due to the same factors that make "multilateral comparison" as practised by Greenberg unconvincing. If we are to learn more about which languages really belong to Hokan and how the family is to be subgrouped, it will be via the program laid out by Haas (1963), namely the acquisition of more and better data, the establishment of phonological correspondences, and the reconstruction of Proto-Hokan and intermediate proto-languages. It is not from the assemblage of lists of vaguely resemblant words that we can expect to improve our knowledge, but from the fieldwork that continues to be done, the study of the vast store of Harrington material, and such initiatives as Kaufman's (1988) reconstruction of Proto-Hokan.

For example, Greenberg (1987) includes Waicuri in Hokan on the basis of a mere six lexical resemblences, Maratino on the basis of two lexical resemblences (Hokan 33 CRY₁, p. 135, Amerind Dictionary 9 ARRIVE, p. 185) and one grammatical resemblence (1st person pronoun, p. 288).

References

- Bright, William. 1954. Some Northern Hokan Relationships: A Preliminary Report.
 Papers from the Symposium on American Indian Linguistics, ed. C. D. Chrétien,
 M. S. Beeler, M. B. Emmenau and M. R. Haas, pp.63-67. Berkeley: University of California Press.
- Brinton, Daniel G. 1891. The American Race: A Linguistic Classification and Ethnographic Description of the Native Tribes of North and South America. Philadelphia.
- Crawford, James M. 1976. A comparison of Chimariko and Yuman. Hokan Studies, ed. Margaret Langdon and Shirley Silver, pp. 177-191. The Hague: Mouton.
- Crawford, Judith G. 1976. Seri and Yuman. Hokan Studies, ed. Margaret Langdon and Shirley Silver, pp. 305-324. The Hague: Mouton.
- Dixon, Roland B. 1905. The Shasta-Achomawi: A New Linguistic Stock, with Four New Dialects. American Anthropologist 7:213-217.
- Dixon, Roland B. 1910. The Chimariko Indians and Language. University of California Publications in American Archaeology and Ethnology 5:293-380.
- Dixon, Roland B. and Alfred L. Kroeber 1913a. Relationships of Indian Languages in California. Science n.s. 37:225.
- Dixon, Roland B. and Alfred L. Kroeber 1913b. New Linguistic Families in California. American Anthropologist 15:647-55.
- Dixon, Roland B. and Alfred L. Kroeber 1919. Linguistic Families of California. University of California Publications in American Archaeology and Ethnography 16:47-118.
- Gatschet, Albert S. 1877. Indian Languages of the Pacific States and Territories. Magazine of American History 1:1.145-171.
- Greenberg, Joseph H. and Morris Swadesh 1953. Jicaque as a Hokan language. IJAL 19:216-22.
- Greenberg, Joseph H. 1987. Language in the Americas. Stanford: Stanford University Press.
- Gursky, Karl-Heinz. 1968. Gulf and Hokan-Subtiaban: New lexical parallels. IJAL 34:21-41.
- Gursky, Karl-Heinz. 1974. Der Hoka-Sprachstamm. Eine Bestandsaufnahme des lexikalischen Beweismaterials. Orbis 23:170-215.
- Gursky, Karl-Heinz. 1988. Der Hoka-sprachstamm: Nachtrag I Abhandlungen der Völkerkundlichen Arbeitsgemeinschaft 58.
- Gursky, Karl-Heinz. 1989. Der Hoka-sprachstamm: Nachtrag II Abhandlungen der Völkerkundlichen Arbeitsgemeinschaft 63.
- Gursky, Karl-Heinz. 1990. Der Hoka-sprachstamm: Nachtrag III Abhandlungen der Völkerkundlichen Arbeitsgemeinschaft 65.

- Haas, Mary R. 1963. Shasta and Proto-Hokan. Language 39:40-59.
- Haas, Mary R. 1964. California Hokan. Studies in Californian Linguistics, ed. William Bright, pp.73-87. Berkeley: University of California Press.
- Jacobsen, William H. Jr. 1958. Washo and Karok: An approach to comparative Hokan. IJAL 24:195-212.
- Jacobsen, William H. Jr. 1979. Hokan inter-branch comparisons. The Languages of Native America: Historical and Comparative Assessment, ed. Lyle Campbell and Marianne Mithun, pp.545-591. Austin: University of Texas Press.
- Jacobsen, William H. Jr. 1986. Washo Linguistic Prehistory Papers from the 1983, 1984, and 1985 Hokan-Penutian Languages Conferences (Occasional Papers on Linguistics, No. 13), ed. James E. Redden, pp. 33-58. Carbondale: Department of Linguistics, Southern Illinois University.
- Kaufman, Terrence 1988. A Research Program for Reconstructing Proto-Hokan: First Gropings. Papers from the 1988 Hokan-Penutian Languages Workshop, ed. Scott DeLancey, pp. 50-168. Eugene, Oregon: Department of Linguistics, University of Oregon. (University of Oregon Papers in Linguistics. Publications of the Center for Amerindian Linguistics and Ethnography 1.)
- Kroeber, Alfred L. 1915. Serian, Tequistlatecan, and Hokan. University of California Publications in American Archaeology and Ethnography 11:279-290.
- Langdon, Margaret. 1974. Comparative Hokan-Coahuiltecan Studies: A Survey and Appraisal. (Janua Linguarum, Series Critica 4). The Hague: Mouton.
- Lehmann, Walter. 1920. Zentral-Amerika. Berlin: Reimer.
- McLendon, Sally. 1964. Northern Hokan (B) and (C): a comparison of Eastern Pomo and Yana. Studies in Californian Linguistics, ed. William Bright, pp. 126-144. Berkeley: University of California Press.
- Olmsted, D. L. 1956. Palaihnihan and Shasta I: Labial Stops. Language 32:1.73-77.
- Olmsted, D. L. 1957. Palaihnihan and Shasta II: Apical Stops. Language 33:2.136-138.
- Olmsted, D. L. 1959. Palaihnihan and Shasta III: Dorsal Stops. Language 35:4.637-644.
- Oltrogge, David F. 1977. Proto Jicaque-Subtiaba-Tequistlateco: a comparative reconstruction. Two Studies in Middle American Comparative Linguistics. Austin: University of Texas Press.
- Rivet, Paul. 1942. Un dialect Hoka Colombien: le Yurumangúi. Journal de la Société des Américanistes de Paris 34:1-59.
- Ruhlen, Merritt. 1987. A Guide to the World's Languages. Volume 1: Classification. Stanford, CA: Stanford University Press.
- Sapir, Edward. 1917a. The position of Yana in the Hokan stock. University of California Publications in American Archaeology and Ethnology 13:1-34.
- Sapir, Edward. 1917b. The Hokan and Coahuiltecan Languages. IJAL 1:280-290.

- Sapir, Edward. 1921. A supplementary note on Salinan and Washo. IJAL 2:68-72.
- Sapir, Edward. 1925. The Hokan affinity of Subtiaba in Nicaragua. American Anthropologist 27:402-435, 491-527.
- Sapir, Edward. 1929. Central and North American Indian languages. Encyclopaedia Britannica. 14th edition. 5:138-141.
- Silver, Shirley. 1964. Shasta and Karok: a binary comparison Studies in Californian Linguistics, ed. William Bright, pp. 170-181. Berkeley: University of California Press.
- Swanton, John. 1915. Linguistic position of the tribes of Southern Texas and Northeastern Mexico. American Anthropologist n.s. 17:17-40.
- Turner, Paul L. 1967. Seri and Chontal (Tequistlateco). IJAL 33:235-239.
- Voegelin, Charles F. and Florence M. Voegelin. 1977. Classification and Index of the World's Languages. New York: Elsevier.
- Waterhouse, Viola Grace. 1976. Another look at Chontal and Hokan. Hokan Studies, ed. Margaret Langdon and Shirley Silver, pp. 325-343. The Hague: Mouton.