SOME REMARKS ON UTO-AZTECAN CLASSIFICATION

MARIO CORTINA-BORJA AND LEOPOLDO VALIÑAS C.

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Introduction. This paper employs a quantitative approach to address a series of qualitative problems posed in the classification of the Uto-Aztecan languages proposed by Miller (1984) based on lexical evidence, on the assumption that the addition of certain quantitative techniques to traditional linguistic analysis could help to refine or resolve certain existing questions and perhaps suggest additional ones. Our intention was not only to enhance Miller's classification with other data analysis methods but to state precisely, i.e., in mathematical form, some of his conclusions.[1]

It is important to stress that Miller's classification (and ours) is based on lexical evidence using Swadesh's (1955) 100-item list with twelve substitutions (cf. Miller 1984:8). Due to the relative ease with which the lexicon can be modified by interlinguistic relations (although Swadesh's list is understood to be somewhat resistant to this kind of influence), this classification would not only reflect language groupings but vicinity relations. In this sense, it is clear that the results here presented have to be compared not only with phonological and morphological evidence, as Miller points out, but also with ethnohistorical and geographical information.

With this in mind, we added six more languages (Southern Tepehuan, Tepecano, and four Aztecs: Nayaritan Mexicanero, Tuxpan [Jalisco], Ostula [Michoacán], and Pochutec)[2] to the cognate table published by Miller (1984:9–12) in order to have a more balanced representation; then we transformed the cognate density table into a lexical distance matrix (one minus the cognate density between pairs of languages) and applied to it some exploratory methods for multivariate data. In other words, we worked with lexical distances among the Uto-Aztecan languages: the

[1] We are most grateful to Barbara Price for her valuable help in the revision of this paper. Special thanks are due to Antonino Guzmán for drawing the graphs. Of course, all the responsibility remains ours.

[2] Southern Tepehuan (Sánchez-Olmedo 1980), Tepecano (Mason 1917, Lumholtz 1904, and Key, n.d.), Nayaritan Mexicanero, Tuxpan and Ostula, Michoacán, Aztecs (Lastra 1986 and Valiñas's field notes), and Pochutec (Boas 1917).

[*IJAL*, vol. 55, no. 2, April 1989, pp. 214–39]

0020-7071/89/5502-0005$01.00

more similar in lexicon are two languages, the closer they are in lexical distance.

The main points of Miller's proposal which concern us are as follows: (*A*) the Northern Uto-Aztecan (NUA) branch; Miller says that "there is no evidence for a Shoshonean or Northern Uto-Aztecan (NUA) grouping (Numic, Tubatulabal, Takic, Hopi)" (1984:13); (*B*) the five major groups: (1) Numic (with three subgroups), (2) Tubatulabal, (3) Takic (with two subgroups), (4) Hopi, and (5) Southern Uto-Aztecan (SUA), which includes two branches, Sonoran (four subgroups) and Aztecan (two subgroups); within this point, the position of Tubatulabal, Hopi, Tubar, and Cora; (*C*) the possible tree representation. Miller points out that their interrelationships "cannot be accurately represented by a family-tree diagram without distortion" (1984:18).

We present our results in graphic form. Our position in this paper is that the distinctive shapes of the graphs themselves constitute evidence for linguistic distance or proximity of subgroups within Uto-Aztecan.

Data analysis methodology. One of the objectives of this paper is to strengthen the classification of the Uto-Aztecan languages in order to use it dynamically, i.e., in a historical way. The data analysis methods employed, each, interestingly, with its own distinctive set of results, were as follows: (*a*) cluster analysis (CLAN), the seven methods currently described in Anderberg (1973) and in Espinosa and López (1977); (*b*) multidimensional scaling (MDS) in its classical and ordinal variations, treated by Shepard, Romney, and Nerlove (1972:1–12) and Gordon (1981); (*c*) nonhierarchical classification methods (K-MEANS), explained by Hartigan (1975); (*d*) Sum of Squares Criterion (SSC) in Edwards and Cavalli-Sforza (1965) and Gordon (1981); and (*e*) descriptive statistics, especially mean, standard deviation, maximum and minimum, coefficient of variation, median, and tests for departure from normality (TDN) (D'Agostino and Pearson 1973).[3] In the application of these techniques we began with the distance matrix, which formed the basic data for the CLAN, the MDS, the TDN, the SSC, and the descriptive statistics; for the K-MEANS we used the coordinates of each language given by the various MDS obtained.

We took advantage of the ANALISIS/CUMULOS program (Espinosa, Reyes, and López 1978) for the computations needed to construct

[3] All the computations involved were performed on the Burroughs B7800 of the Universidad Nacional Autónoma de México (UNAM) and on the microcomputers of the Instituto de Investigaciones Antropológicas, UNAM. We are indebted to the Dirección de Cómputo Académico, UNAM for the computer mainframe resources used.

and evaluate the dendrograms, and of the routines implemented by Iraizos (1984) to generate MDS configurations, and by Ardisson (1980) for the K-MEANS.[4]

ANALISIS/CUMULOS allows us to obtain seven trees or dendrograms differentiated according to the particular way of calculating the distances between the groups formed.[5] The potential selection of one of these HIERARCHICAL trees can then be achieved in several ways. We had two basic options: (*a*) to select the most adequate dendrograms based on the similarities to the more accepted linguistic classifications; or (*b*) to use data analysis criteria (e.g., MDS, K-MEANS, and SSC) which are independent of the CLAN methods; these criteria would allow us to assess measures of the internal homogeneity of the groups generated, and thus permit us to minimize the internal dispersion within the groups. In an attempt to exploit all the information contained in the data, we chose the second option.

Using classical MDS we obtained "maps" of the languages in two and three dimensions. This kind of MDS searches a configuration of points in a space with a certain number of dimensions. This is attained by using projections over some axis in order to concentrate (into a set of coordinates) the dispersion between languages which is contained in the distance matrix. The evaluation of the goodness of fit for the configuration obtained is made in terms of the proportion of dispersion represented on each axis. This figure is additive and decreases (for every axis) in terms of the number of dimensions. This means that the dimensions are INDEPENDENT and that the first ones reflect more information about the data. The goodness of fit proved to be satisfactory: working with three dimensions, we always included more than 91 percent of explained variation.[6]

In the ordinal MDS configurations in two and three dimensions, the components are obtained jointly and thus lack specific weight. A goodness-of-fit index for the Euclidean distance matrix produced with

[4] All the computer programs not mentioned in specific references were developed by Cortina-Borja, such as those for evaluating dendrograms, calculating descriptive statistics, preparing raw data for further analysis, etc.

[5] If readers are interested in receiving specifics about all the statistical methods employed here, about the computer programs, or about the six languages added, we would be happy to mail additional information (send requests to us at the Instituto de Investigaciones Antropológicas, U.N.A.M., C.U., Mexico, D.F. 04510, Mexico). However, Shepard et al. (1972), Gordon (1981), and Martínez Malo (1979) have very good and accessible introductions.

[6] The first component contains 65.24 percent of the total dispersion, the second, 19.83 percent, and the third, 6.62 percent (91.69 percent together).

the points of the configuration is called STRESS. It accounts for differences between this matrix and the original data. If—as occurred with our ordinal MDS solutions[7]—its value is less than 10 percent (or 0.1), it can be considered sufficient.

The K-MEANS method produces nonhierarchical classifications. It acts the inverse way of CLAN: given the number of groups desired (K), it gives the partition of the items in such a way as to minimize the internal variance of the groups; each group is focused upon a proposed CENTROID and items are seen as belonging to such a defined group on the basis of their nearness to a centroid.

The SSC method was used as a form of evaluating the internal homogeneity of some groups proposed on the basis of results obtained with the other methods. The SSC produces an internal variance measure calculated from the contribution of all the groups to total variation. It should be clear that this figure decreases as the number of groups increases. Although it would have been possible to seek for THE partition which attains the maximal homogeneity for a range of number of groups, this is impossible in practice because of the enormous amount of computer time required to search for all the possible partitions. Instead, we proposed some potential partitions and developed our argument on the basis of the best, the one which had the smallest SSC index.

The descriptive statistics and TDN were calculated over the columns of the distance matrix. The first ones were used in three ways. First, they provided useful information in a number of steps of the analysis. Second, plotting the median versus the standard deviation offered a method of assessing the isolation of some languages. Those located consistently distant from the majority of the languages would appear toward the upper-left corner of the scattergram. Conversely, the upper-right corner would contain the languages with a lot of variation in their distances but with a predominance of long distances: these languages would form steady groups. Third, plotting the distances to a language versus the distances to its nearest neighbor and to some distant languages provided a device for seeing the relationships contained in the distance matrix. We expected that these graphs would show the groups in a straightforward way.

Finally, the TDN were used as an attempt to obtain a more precise idea of the frequency distribution of the distances for each language. First, if such a distribution were asymmetrical, the mean (average) would not be the best measure of location; the median would be better.

[7] Gordon (1981) says that a reasonably good fit should have about 0.1 of stress. With three dimensions we got 0.07; with two, 0.13.

Its use was supported by the results obtained with the TDN, which pointed to a clear nonnormal distribution in the majority of languages. Second, if a distribution had a nonsignificant deviation from a normal distribution, this fact could be used as evidence of the isolation of a language: the nonnormal distributions of distances should have a skew only as a long tail toward the lower limit of the range of distances and must overall be narrower than the normal.[8] If these conditions are not met, the language should have a greater proportion of long distances than the biased-and-narrower-distribution languages (i.e., languages which belong to a tight group).

Cluster analysis and configuration of minimum sum of squares. As we noted before, with this method we obtained seven different dendrograms. The results showed three different situations: (*a*) in only three dendrograms (the first, second, and sixth) the first major division was between Northern Uto-Aztecan (NUA) and Southern Uto-Aztecan (SUA) groups, although if we cut them at the point where each showed five and six defined branches, no single one represented satisfactorily the generally accepted classification based on linguistic arguments.

The cuts gave the following five groups (the sixth group was obtained by the branching of the one in parentheses): *First tree*: 1. Hopi, 2. (Gabrielino, Takic), 3. Tubatulabal-Numic, 4. Sonoran, 5. Aztecan (note that this one rather resembles the one proposed by Miller 1984:18, in his discussion of the Takic pruning). *Second tree*: 1. Hopi-Tubatulabal, 2. Takic, 3. Numic, 4. (Corachol, Sonoran), 5. Aztecan. *Sixth tree*: 1. Hopi-Tubatulabal-Serrano-Gabrielino, 2. other Takic, 3. Numic, 4. Tepiman-Tarahumaric-Opatan, 5. (other Sonoran, Aztecan).

(*b*) In three other trees, we find such division but with a "meaningless" mixture between both branches (where one or two Northern languages appeared among the Southern ones): in the third, Hopi was inside the Sonoran languages but splitting only shortly after Aztecan (the first partition inside SUA was Aztecan/Sonoran, and the first within the Sonoran was Hopi/Sonoran); in the fifth, Hopi was not only deep inside the Sonoran branch, but it appeared as a variant of Tepiman; and in the fourth (the oddest one), Serrano and Gabrielino were inside the Southern group, Serrano as part of the Sonoran languages, Gabrielino as part of Aztecan.

(*c*) Only one dendrogram (the seventh) did not show as its first major division the one between NUA and SUA. Instead, we found that the first

[8] A normal distribution (also called Gaussian) follows a bell-shaped curve.

group that branched was Numic, then, from the rest of Uto-Aztecan, Aztec, and then Takic (including Tubatulabal and Hopi).

When we used SSC to evaluate the dendrograms, the tree that seemed to be the most adequate was the second one,[9] at least in its first five partitions (figure 1).[10] Although this dendrogram is not the most similar to Miller's classification, it shows some interesting facts; for example, after the initial NUA-SUA separation, SUA is the first to undergo an internal split: Aztecan separates from Sonoran.

On the other hand, the first division inside the NUA branch occurs between the Numic languages and the others. Then the Hopi-Tubatulabal group splits from the Takic branch, which later shows an initial division between the Serrano-Gabrielino and the Cupan, before Hopi and Tubatulabal, in their own branch, are well defined. This tree (without considering the place of Hopi) is similar to Miller's (1984:18) treatment of the Numic pruning.

Several consequences follow from this discussion: (*a*) if we consider the SUA branch as a distinct branch, we should consider as comparable in status the NUA, because Aztecan split from Sonoran before any separation within the NUA branch. (*b*) Tubar seems to be a subgroup of Cahitan; (*c*) there seems to be a linkage of Hopi and Tubatulabal and both with (inside?) Takic (through Gabrielino?); (*d*) the division of Takic into two subgroups (Serrano-Gabrielino and Cupan, perhaps with Luiseño as a transitional language) is ephemeral, because the component subgroups begin to diversify almost immediately. We return to these questions later.

We also applied the SSC to the branches. The "best" tree representation of the NUA branch (the one with the smallest variance, i.e., the one that made the most homogeneous cuts) is the one shown in figure 1.

[9] This one was obtained by the so-called longest link method; after the two languages that have the smallest distance coefficient are joined in a new group, the whole distance matrix is recalculated considering now the biggest distance coefficient between one of the members of the newly formed group and the other groups or languages. Then, again, the smallest coefficient joins two members in a group and the recalculation process is repeated. The whole process is repeated until all the languages form a unique group.

[10] Mono (Mo), Northern Paiute (PN), Panamint (Pa), Shoshoni (Sh), Comanche (Cm), Southern Paiute (PS), Ute (Ut), Chemehuevi (Ch), Kawaiisu (Ka), Tubatulabal (Tb), Serrano (Se), Gabrielino (Ga), Cupeño (Cu), Cahuilla (Ca), Luiseño (Lu), Hopi (Ho), Papago (Pg), Nevome (Ne), Northern Tepehuan (TN), Southern Tepehuan (TS), Tepecano (Tc), Tubar (Tu), Opata (Op), Eudeve (Eu), Tarahumara (Ta), Guarijío (Gu), Yaqui (Yq), Mayo (My), Cora (Co), Huichol (Hu), Classical Aztec (Nc), Tetelcingo (Te), Zacapoaxtla (Za), Nayaritan Mexicanero (Mx), Tuxpan (Tx), Michoacán (Mi), Pipil (Pi), and Pochutec (Po).

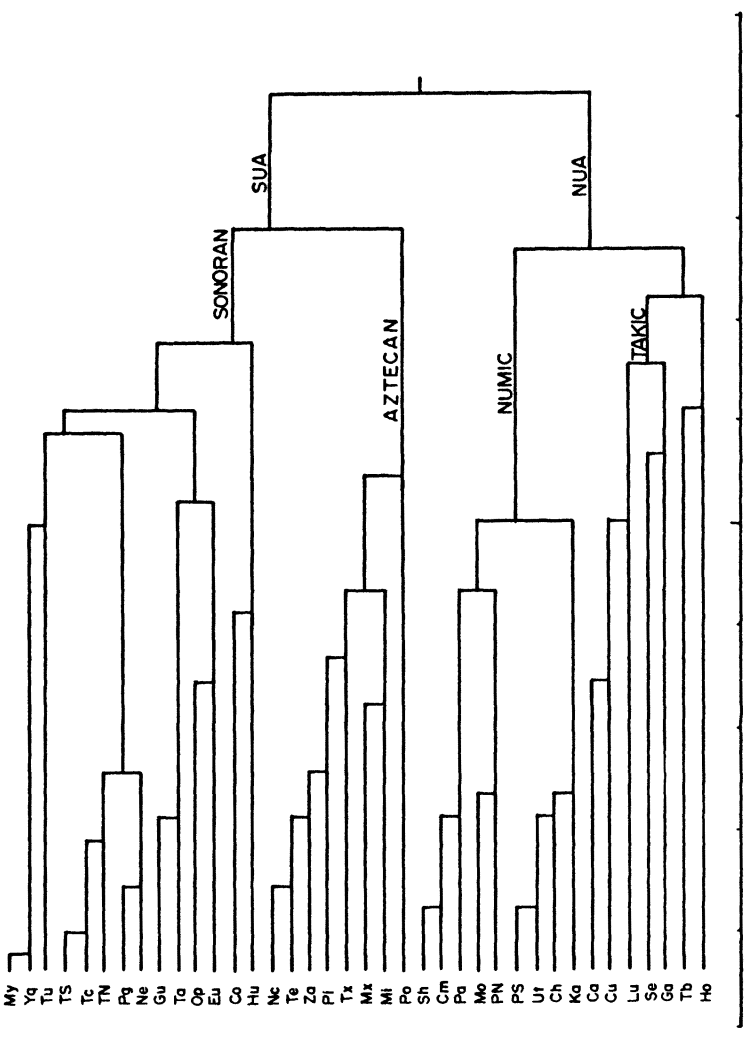

FIG. 1.—Uto-Aztecan tree.

About the SUA branch, it is interesting to point out that no one of the seven trees obtained by the ANALISIS/CUMULOS program represented the whole SUA-branching as the "best" one (as it was evaluated by SSC). The first groups of the SUA to split were Aztecan, Corachol, and Tepiman. Tubar appeared linked with Cahitan, splitting after Tarahumaran and Opatan. Aztecan shows two main groups, Pochutec and General Aztec. This latter has two clear groups: (*a*) Western (Nayaritan-Tuxpan-Michoacán) and (*b*) Central (Classical Aztec-Tetelcingo-Zacapoaxtla-Pipil).

At this level, we had a hierarchical clustering. However, this is not enough to explain the classification: it did not account for some facts (for instance, the relationship between Cahuilla and the other Takic, and the fact not shown by any tree that Serrano is lexically closer to Cahuilla than to any other language).

Multidimensional scaling and K-MEANS. In the interest of greater accuracy, we worked with two types of MDS, the classical and the ordinal.

Classical MDS—in the classical configuration it is easy to assume that there are four groups (as we can see in figure 2 if we take only the first two dimensions): Numic, Takic-Hopi-Tubatulabal, Sonoran, and Aztecan. The position of Tubatulabal and Hopi is very special; it is hard to say whether they are transitional between Takic and Numic. Tubatulabal is "between" Takic and Numic, but it is clearly closer to Takic. Hopi is also closer to Takic. Again, in this "map" we can say that it is possible and reasonable to talk about the two major subgroupings of the Uto-Aztecan languages.

If we look at the three dimensions, we can see in detail the internal relationships of those groups. The degree of diversification between groups is indicated on the graph by the distribution of heights of component points (languages). Now it is more evident that there are five groups (considering Hopi and Tubatulabal as comprising a distinct group because of their positions on the graph). On the basis of the first component (Y-axis) we can see that the groups are internally very tight. It is important to say that the Y-axis is the most significant and the Z-axis the least meaningful of the three.

It is interesting to note that an important step in this kind of MDS is to relate the components with the information contained in the data. In this sense, in figure 2 the first component (Y-axis) may be associated with a geographical distribution: there is, in general terms, a north-to-south distribution. It was very hard to associate the second and third components (X and Z axes) with some consistent kind of relationships among the data. We leave this point to further analysis.

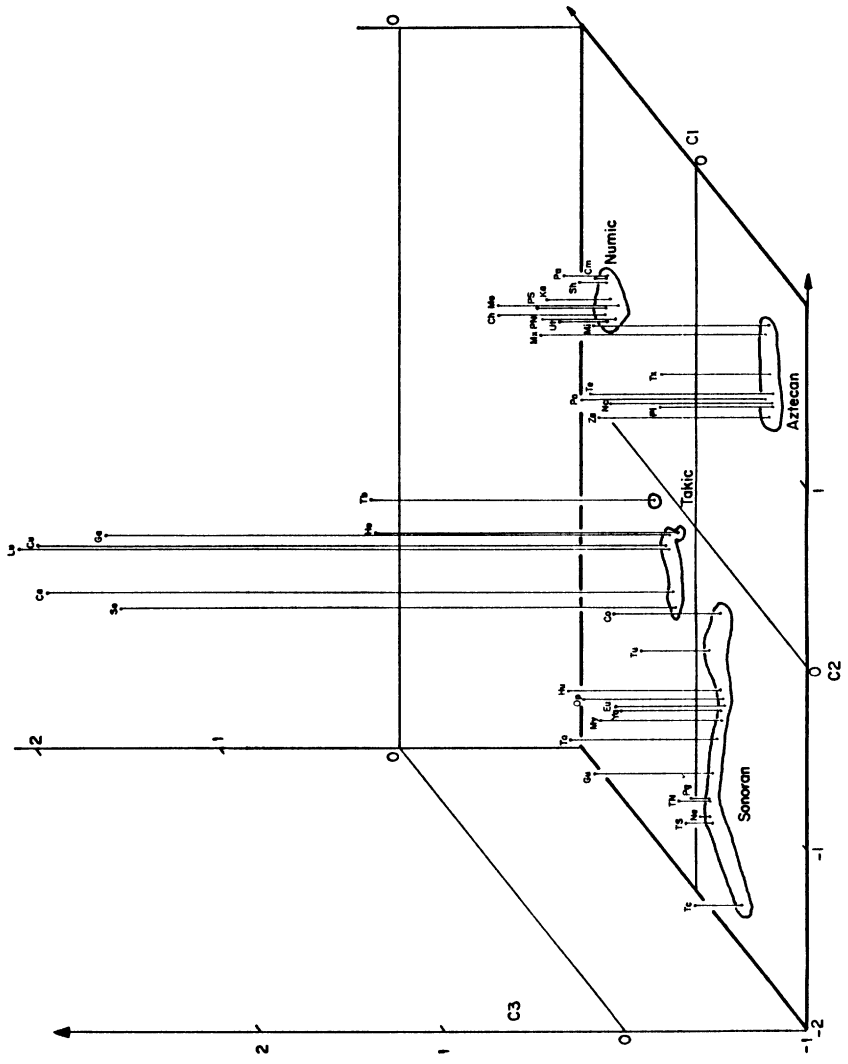

FIG. 2.—Classical MDS (three-dimensional).

If we look at the three components jointly, we can distinguish the following groups:

(*a*) Numic. Three subgroups are more or less clearly defined. Nevertheless, Kawaiisu seems somehow transitional; it is intermediate in terms of Northern Paiute, Southern Paiute, and Panamint. Central Numic shows more internal coherence.

(*b*) Takic. The two possible subgroups are not defined with real clarity; however, Serrano and Gabrielino are rather more distant from the other Cupan.

(*c*) Hopi and Tubatulabal. In some sense, these can be taken as a single group, and analyzed additionally as transitional languages: Hopi between Takic and Sonoran, and Tubatulabal between Takic and Numic. Note the differential height on the graph (figure 2): at this point, in other words, the branching is directly observed.

(*d*) Sonoran. These languages appear as a spread group, with Tepiman as somewhat more differentiated from the other Sonoran and constituting a tight group. Tepecano is truly separated from the rest (but still inside Tepiman). Cora and Tubar could be considered as two subbranches, although they are quite similar to Taracahitan. Huichol appears more closely linked to Taracahitan than to Cora. With this map, it might be better to point out Sonoran as a mesh.[11]

(*e*) Aztec. This group seems to be very homogeneous, although within it Mexicanero-Michoacán and Tuxpan constitute apparent poles.

Ordinal MDS—figure 3 presents the 2-D configuration of the ordinal scaling. Again, it is more or less clear that we can talk about four groups: Aztecan, Sonoran, Takic-Hopi-Tubatulabal, and Numic. This figure very clearly suggests the potentially transitional status of Hopi and Tubatulabal. It is also possible to establish only two major partitions or supergroups: NUA-SUA. It may be significant to note that, again, while Hopi and Tubatulabal show close ties to each other, Hopi's nearest neighbors are Takic, and Tubatulabal's are Numic.

In terms of distances between groups or isolated languages, the following facts are evident:

(*a*) The shortest distance between any Aztecan language and any Sonoran (Zacapoaxtla-Tepecano) is clearly longer than any distance within the Aztecan group (Nayarit-Zacapoaxtla). And, if Cora is excepted, the same could be said about any distance within the Sonoran group (Tubar-Tepecano is the longer).

[11] In two other papers (Valiñas and Cortina, forthcoming; in preparation) we worked only with this branch. Because of the other languages and distances, the internal diversity is somewhat masked.

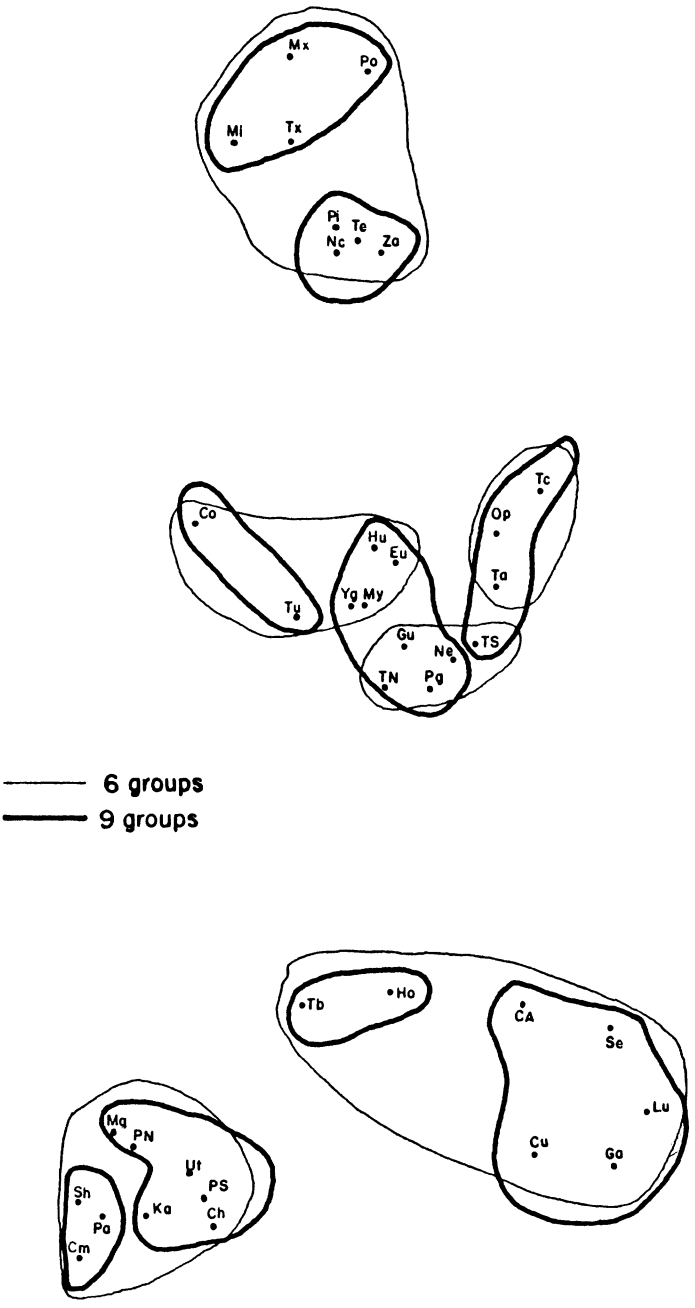

FIG. 3.—Ordinal MDS (two-dimensional), showing K-MEANS configurations.

(*b*) Cora is closer to Central Aztecan (Tetelcingo, Zacapoaxtla, Classical Aztec, Pipil) than to Tepecano.[12] This relationship may be explicable in part by the fact that Tepecano has many zero forms, i.e., much missing data.

(*c*) Again, excepting Cora, the shortest distance of any Takic language or Hopi or Tubatulabal to any Sonoran (Papago-Cahuilla) is longer than any distance within the Sonoran group.

(*d*) The shortest distance between any Numic language and Tubatulabal, Hopi, or any Takic is longer than any distance inside the Numic group (Comanche-Southern Numic is the longest).

(*e*) The nearest language to Hopi is Tubatulabal, and vice versa.

(*f*) The shortest distance between Hopi and any Numic (Ute) is shorter than that between Hopi and its most distant Takic languages (Gabrielino and Luiseño).

(*g*) The distances between Tubatulabal and its most distant Numic (Comanche), and its closer Sonoran (Papago and Northern Tepehuan), are smaller than those between Tubatulabal and some Takic (Luiseño or Gabrielino).

(*h*) Hopi is closer to any Tepiman than to any Central Numic.

(*i*) The smallest distance between Aztecan and Sonoran is shorter than the shortest distance between Sonoran and any NUA.

(*j*) The longest distance between Numic and Takic (Comanche and Luiseño) is shorter than the longest distance inside SUA (Mexicanero and Papago).

With all these facts we can propose the following:

(*a*) There is a clear division between NUA and SUA.

(*b*) Hopi and Tubatulabal can both be considered as transitional languages. It is possible to say that they allow the linkage of Numic with Takic.

(*c*) Numic shows its three subgroups: Central, Southern, and Western, with Kawaiisu located almost equidistant among them.

(*d*) Takic is a spread group; the distances within it could permit us to consider Hopi as a Takic language.

(*e*) Sonoran is a mesh.

(*f*) There are two clear groups within Aztecan.[13]

[12] For convenience, we refer to Classical Aztec, Tetelcingo, Zacapoaxtla, and Pipil as Central. Nayaritan Mexicanero, Tuxpan, and Michoacán are Western. These labels are only approximations of the facts of dialectology (see Lastra 1986:190).

[13] This division corresponds to the languages we added to Miller's original sample. This could be due to the many zero forms (treated as noncognate items) extant in the Aztec added. This presence of zero forms is a regional fact (see Valiñas 1981*b*).

In the interest of refining our accuracy, we applied the K-MEANS method to this configuration and obtained the following results (see figure 3):

(*a*) Numic first separated into two groups (Central and Western-Southern), with Western as a kind of transitional. Also, before forming its own group, Northern Paiute clustered with Southern, and Mono with Central.

(*b*) Tubatulabal and Hopi formed a group by themselves; later they were associated with Takic.

(*c*) Takic appears as a unique group; in the process of analysis, the link with Tubatulabal and Hopi becomes evident, but only at a higher level of clustering.

(*d*) Sonoran showed no consistent internal clustering. Sometimes we obtained two different three-group formations: Cora-Tubar, Opata-Tarahumara-Tepecano-Southern Tepehuan, and the other Sonoran; and Corachol-Cahitan-Tubar-Eudeve, Tepiman (without Tepecano), and Tarahumara-Opata-Tepecano. But also we got two groups: Corachol-Tubar-Cahitan-Eudeve-Northern Tepehuan and the other Sonoran. This could mean that Sonoran is a mesh.

(*e*) Aztecan showed two groups (Western-Pochutec and Central) and became well defined approximately at the time of the emergence of Numic.

It is noteworthy that when we forced the generation of three clusters, the result was 1. Numic-Tubatulabal (!), 2. Hopi-Takic-Sonoran (without Cora) (!), and 3. Cora-Aztecan (!). As expected, with a reduction to two groups, the NUA-SUA division appeared.

In figure 4, we see the three-dimensional map of the ordinal MDS. This result is more accurate than the two-dimensional one because stress is reduced. The relevant facts are as follows:

(*a*) It is clear that there are two major groups, the NUA and the SUA. The first is upward (with positive values in the third dimension), the latter is downward (with negative values). Figure 4, as a result, shows a distinct step in its graph, an operationalized statement of the validity of this grouping.

(*b*) Within NUA, we can see the four generally accepted groups.

(*c*) Within Numic, again Kawaiisu appears as transitional among the three subgroups.

(*d*) Tubatulabal is almost halfway between Takic and Numic, but somewhat nearer to Numic.

(*e*) Takic shows three subgroups: Cupeño-Luiseño, Serrano-Cahuilla, and Gabrielino, with this last the most distinct of all the Takic languages.

(*f*) Hopi appears strikingly isolated, if still in the NUA branch. It is closer to Tubatulabal than to any other (their distance is shorter than

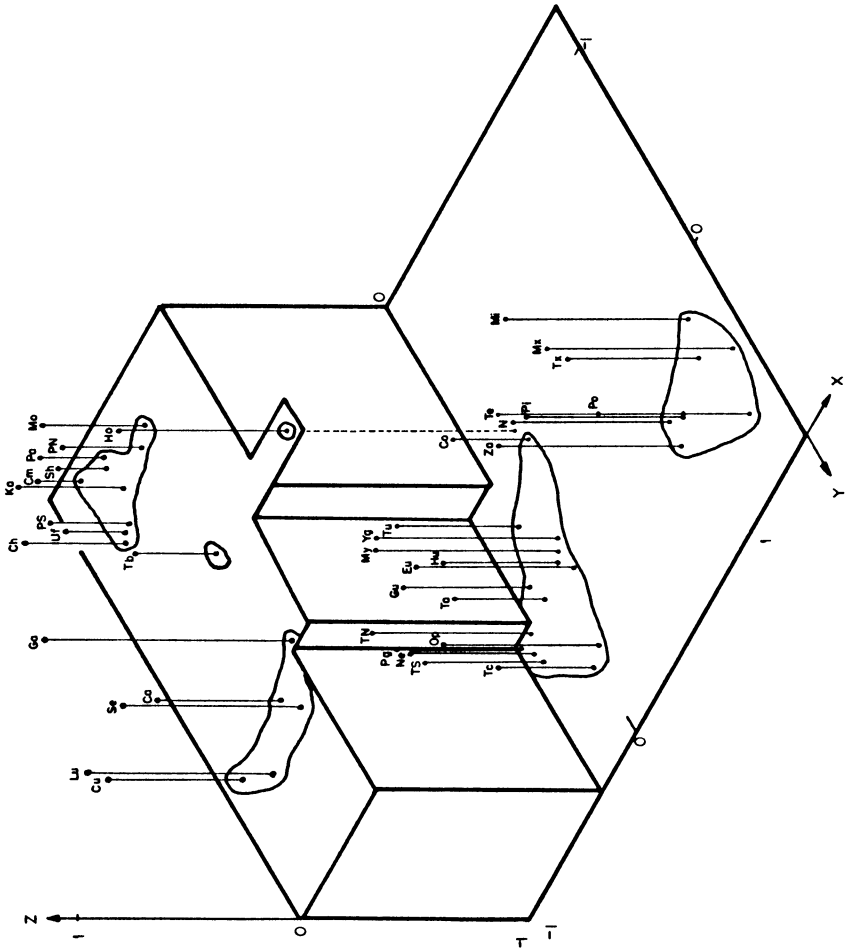

FIG. 4.—Ordinal MDS (three-dimensional).

that between Gabrielino and any Cupan). In some sense, it is near Takic, but it is closer to Cahitan than to Kawaiisu or Luiseño.

(*g*) Sonoran still appears as a mesh. But now we can possibly identify some subbranching: for instance, Tepiman, Cahitan, and also Tarahumaran. It is hard to say, instead, that Opatan or Corachol are more or less well defined subgroups.

(*h*) Cora is relatively isolated, but closer to Tubar.

(*i*) Aztecan shows two subgroups, Central and Western-Pochutec.

When we applied K-MEANS to the 3-D MDS configuration, the results differed somewhat from those derived from the 2-D MDS, but in general terms the resemblance was close. The interesting facts were the following:

(*a*) Numic formed a clear group only after Takic was clearly defined, and it was always joined with Tubatulabal. The Southern and Central Numic were unified in one subgroup.

(*b*) Tubatulabal never appeared alone, but always with some or all Numic languages (always with Western, and never with Central Numic).

(*c*) Takic either constituted a group by itself, or appeared with Hopi.

(*d*) Hopi appeared alone or with Takic.

(*e*) Sonoran never showed any consistency, but was always capable of many different alternative subdivisions. On two different occasions it showed two branches, but dissimilar ones. The first attempt yielded Cahitan-Opatan-Huichol-Tepecano on the one hand, and Sonoran on the other; the second resulted in a division between Opatan-Cahitan-Tepecano and Sonoran respectively. In another cluster Sonoran formed three branches: Tepiman (without Tepecano), Opatan-Cahitan-Tepecano, and Tarahumaran-Tubar-Corachol. When we asked for five groups for all the Uto-Aztecan, Sonoran split still differently: Cora-Tubar and the other Sonoran. Only when six and four groups were requested did all the Sonoran languages appear together as a single group.

(*f*) Aztecan emerged early as a distinctive group; it was in fact the earliest group formed. Its only internal division was between Central and Western-Pochutec.

It is worth nothing that when we sorted for three groups, they were unified as follows: 1. Numic-Tubatulabal, 2. Takic-Hopi-Sonoran, and 3. Aztecan. And, again, with two groups requested, NUA and SUA were defined. This constitutes the fourth demonstration of the validity of this overall partition.

Descriptive statistics. Figure 5 shows the plotting of the medians (MD) of the distances versus the standard deviations (SD) of the distances. The SD might be regarded as a measure of degree of grouping:

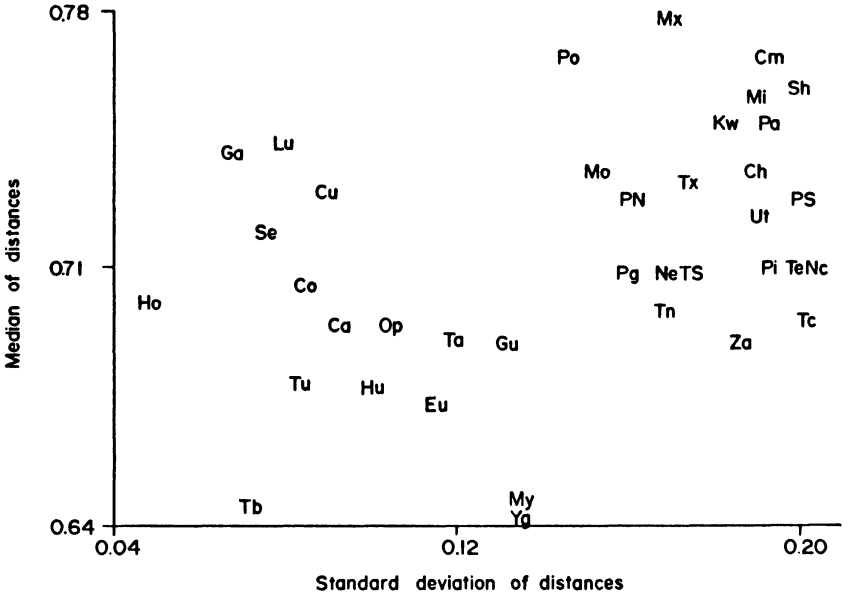

FIG. 5.—Centrality vs. individuality.

the lower it is, the higher the degree of individuality. On the other hand, the MD reflects the degree of centrality with respect to the other languages: the lower it is, the stronger the centrality. In this sense, the languages that are toward the upper-left corner will be those that are most consistently distant from the rest, with the greater tendency to isolation. The languages toward the lower-left corner are truly isolated but with nearly constant small distances to the rest of the languages. Conversely, the languages located close to the lower-right extreme constitute parts of groups with a great internal coherence and centrality. Finally, those at the upper-right corner belong to well-formed groups which themselves manifest a higher degree of isolation as groups.

The reader should note that figure 5 represents only degree of grouping, unrelated to linguistic proximity. This means that if we add to the distance matrix an additional language which is closer to one already plotted, the position of the latter in the graph should be modified toward the lower-right corner. Similarly, if a language is deleted from the distance matrix, its nearest neighbor should move in the upper-left direction.

Figure 5 therefore displays the following facts:

The graph showed a clear-cut division into two main groups: Hopi-Takic-Tubatulabal-Sonoran (without Tepiman) and Aztecan-Tepiman-Numic. The former comprises families that are not so tight as those of the latter.

Numic—it forms a tight noncentral group. The clearest division is bipartite into Western and Southern-Central.

Tubatulabal—appears as one of the most central languages. Its strong centrality indicates its linguistic isolation.

Hopi—it has the greatest degree of individuality, with medium-termed distances.

Takic—as can be seen, each of its components may be treated as a group by itself. Their great separation from the others gives them the appearance of a whole group.

Sonoran—Tepiman is a tight subgroup and somewhat distant from the rest of the family. Cahitan is the most central subgroup (e.g., Yaqui and Mayo are the second and third nearest neighbors to Hopi in the distance matrix, a fact which is not apparent in the graph). The rest of Sonoran represents rather diffuse, poorly formed subgroups. It is worth noting that the roles of Cora and Tubar within Sonoran are analogous to those played by Hopi and Tubatulabal, respectively, regarding the Uto-Aztecan family as a whole.

Aztecan—it is a tight group; nevertheless, the Central subgroup is more central than the Western one. The leftward position of Pochutec suggests that it could be considered a distinct subgroup within Aztecan.

When we analyzed the distances to pairs of neighbor languages, we found a strong direct linear correlation in all the plottings (figure 6 shows the Shoshone-Comanche graph as an example). But when we selected some pairs of very distant languages, the results were parabolic decreasing relationships. This effect is caused by the presence of languages very distant from those represented on the axes (viz. figure 7, the Comanche-Pochutec graph; compare this figure to figure 6: the closer the languages of the axes are, the straighter the line).

Our graphic representations thus reveal the following:

(*a*) The Uto-Aztecan family can be analyzed as a continuum of groups, with Numic and Aztecan as its respective extremes. However, this continuum is not linear: Sonoran and Takic are almost at the same distance to any pair of Numic languages (as is shown in figure 6).

(*b*) Hopi and Tubatulabal could be interpreted as transitional languages because the graph shows no linear correlation (figure 8; cf. also figures 6, 7, and 9).

(*c*) The Takic plottings show Sonoran and Numic as almost equidistant.

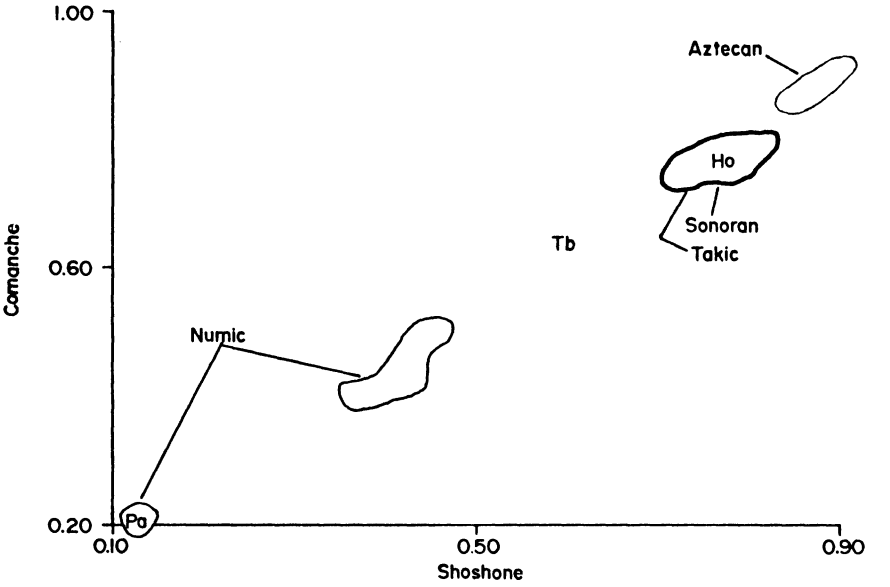

FIG. 6.—Comanche–Shoshone distances.

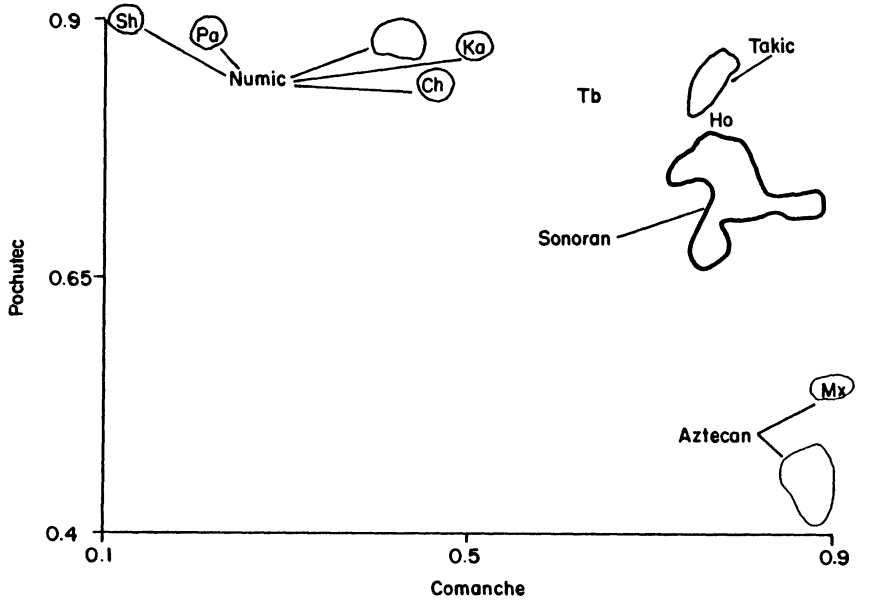

FIG. 7.—Pochutec–Comanche distances.

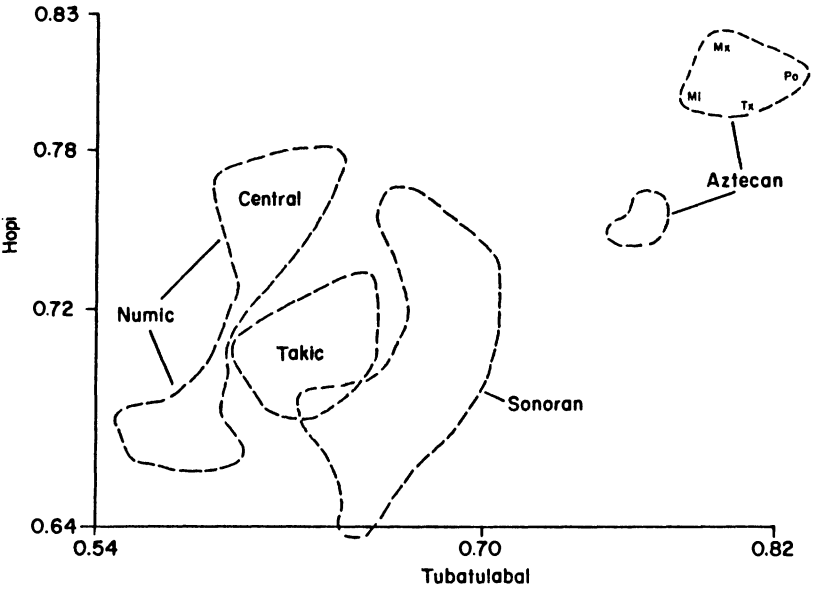

FIG. 8.—Hopi–Tubatulabal distances.

(*d*) Sonoran can be seen as the most central branch: the plots show that languages of relatively distant groups (e.g., Takic, Numic, and Aztec) are in fact approximately equidistant from Sonoran. Figure 9 illustrates this fact with the Guarijio-Tarahumara plotting.

(*e*) Aztecan is shown as the group most distant from any other. It was the only Uto-Aztecan branch that never mixed groups in the paired-nearest-neighbors graphs and consistently yielded the order Sonoran-Takic-Numic.

With TDN we found that only Pochutec, Opatan, Corachol, Tubar, Hopi, Tubatulabal, and Gabrielino-Serrano could be regarded as having normally distributed distances. In other words, they share similar proportions of lower and higher distances. On the other hand, Cupan, Cahitan, and Tarahumara showed strikingly peaked distributions; the reason is that many of their distances are concentrated near a value smaller than the median, combined with a long right tail formed with very few distances. The rest of the languages have skewed distributions but with peaks not as pronounced as for those mentioned above.

It is noteworthy that Tubatulabal, Hopi, Tubar, and Cora again behave similarly when graphed. Also it is interesting to point out that

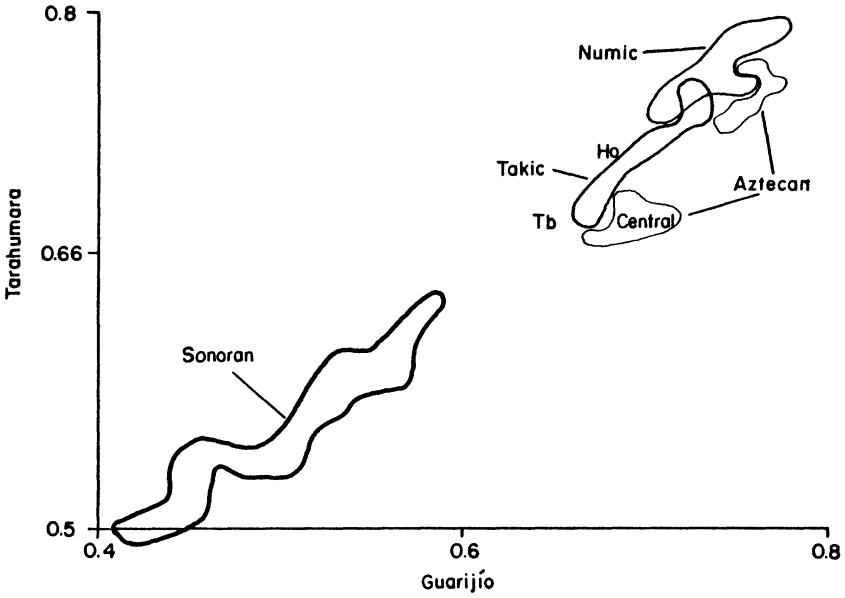

FIG. 9.—Tarahumara-Guarijío distances.

Pochutec and Gabrielino may be comparable: they are the most distant members of their own groups.

Discussion. Much, of course, remains to be explained. Our intention has been to reinforce four points in Miller's proposed classification. Our conclusions at this point of the analysis are as follows:

(*A*) We think that there are many lines of evidence that allow us to justify a NUA-SUA partition. All the methods used showed this division; if we are able to talk about a SUA branch, then we have to consider a NUA branch too. The lexical diversity between Sonoran and Aztecan is greater than that between Numic and Takic; if, therefore, SUA is incontrovertibly a group, by the same criteria, so too is NUA. Hopi and Tubatulabal appear as the nexus, the point of branching between Numic and Takic within NUA.

(*B*) At least lexically, all the methods we employed show the following classification:

Within NUA, we can talk of four groups (or perhaps three):

Numic. It comprises three branches. Miller points out that his data "do not give evidence for any special subgrouping, nor for Central

Numic forming a link between the other two" (1984:16). When we represented their relationships by a tree, we found that the clearly hierarchical condition strongly indicated a special relation between Central and Western Numic. Nevertheless, there is no clear support to this linkage. What is really clear is that these branches form a kind of triangle, with Kawaiisu as its center, but still most closely related to Southern Numic (see figure 3).

Takic. There is no doubt that this group is not a tight one, but it still can be considered as a group because its internal cohesion is maintained on the basis of the distances separating Takic from all other groups. Working with MDS, we found that it was not entirely true that there was "more internal diversity within Takic than within all of Sonoran" (Miller 1984:17). We suggest, if cautiously, that it would be preferable to see Takic as composed of five branches rather than analyzing it in terms of two principal branches. If, for example, the analysis of the internal branching of Sonoran is taken as a model, we note that internal distances within accepted branches, in Sonoran, are far closer than is the case within Takic. It follows that if we analyze Sonoran, therefore, as a mesh, there is at least equal justification for adopting this same strategy for Takic. If, in fact, there were more surviving evidence from more Takic languages, we suggest that the graphic configurations of this group would have resembled those of Sonoran even more closely.

If we look at the MDS maps, we can see at least three facts that allow us to propose a five-branched structure of Takic: (*a*) Gabrielino and Serrano appear close to each other in Classical (figure 2) but not in Ordinal 2-D (figure 3); (*b*) Luiseño is closer to Cupeño than Cupeño is to Cahuilla; and (*c*) Cahuilla seems to be equidistant from all other languages within Takic (it is also the most central of them, as can be seen in figure 5).

This strongly implies (see above) that a tree configuration is insufficient to explain the Takic classification, as is evident if we observe the nearest neighbor of each Takic language: although Serrano's nearest neighbor is Cahuilla, Cahuilla's is Cupeño. Note that Serrano, Cupeño, and Luiseño all share Cahuilla as their nearest neighbor.

Lamentably, we lack the additional Takic languages or dialects that would allow us to reinforce—or, alternatively, to falsify—our assumption; at this point, we consider that the data tend to support the mesh model rather than the tree model.

Tubatulabal. The first three figures clearly highlight the close relation between Tubatulabal and Hopi. Reference to the Classical MDS (figure 2) indicates that after its relationship with Hopi, Tubatulabal is situated

midway between Chemehuevi and Gabrielino; the graphic distances, however, show clearly that its closest proximity is to Western and Southern Numic. Similarly, in the Ordinals (figures 3 and 4) its closest relationships are with Numic (Ute and Southern Paiute). Taken together, these observations indicate the "intermediate" position of this language.

At the same time, of course, Tubatulabal can be considered an isolated language, although it could be linked with Hopi within a NUA branch because of their strong relationship. If this is true, and since both are, in terms of the definitions we have presented, very central, we can say that they separated very early from the "main NUA branch."

We propose Tubatulabal as a transitional language between Numic and Takic but closer to Numic (see figure 8). If there is a linkage with Takic (as the dendrogram suggests), this is due rather to its resemblance to Hopi (which is the most central) and not the consequence of the specific linguistic characteristics of Tubatulabal: the GRAPHIC position of Hopi alters the relative position of Tubatulabal (see especially figures 1 and 3).

At this point we agree with Miller when he says that Tubatulabal has "the sort of link typical of the mesh principle," but not when he points out that there is no evidence "of a closer relationship with either a Numic or a Takic subbranch" (1984:16). As is shown in our graphs, there is a distant but special relationship, particularly with Southern Numic.

Hopi. As has been noted above, it is possible to see Hopi and Tubatulabal as constituting a NUA subbranch with a very early separation from the main NUA branch and from each other. This results from the centrality of both and their intermediate graphic location (figures 2, 3, and 4).

This lexical approach suggests that it is hard to sustain the traditional view of Hopi as a language transitional between Takic and Sonoran (figure 4). Although its nearest neighbors after Tubatulabal—as can be seen in figure 8—are Sonoran (Yaqui and Mayo), additional facts do not permit support of that assumption of transitional status: (*a*) the shortest distance between any NUA language and any SUA is as great as that between Tubatulabal and Northern Tepehuan; (*b*) many Sonoran languages are closer to Tubatulabal than to Hopi (see, for example, figure 9); and (*c*) Hopi is the most central language within all of Uto-Aztecan; i.e., it is virtually equidistant from nearly all the Uto-Aztecan languages. If, therefore, we treat Hopi as transitional between north and south, then logically we should apply the same argument to Tubatulabal, but if so, the result is contrary to fact.

All these observations suggest to us that NUA is most clearly analyzed as a mesh; as Miller pointed out, a tree classification is insufficient as a representation of the internal interrelationships within NUA.

Within SUA, on the other hand, we can talk of only two well-defined groups:

Sonoran. There is no doubt that Sonoran is a mesh, but its internal subgrouping is nonetheless clear, albeit with two remaining points of discussion: (*a*) Corachol and Tepiman are, as Miller pointed out, "the two extremes within the Sonoran subfamily" (1984:17). Although Cora is the most isolated language within the branch, it "is linked to the other Sonoran languages through Huichol" (1984:17). In this sense, if there were no Cora languages, Huichol could be classified with Taracahitan on the basis strictly of the internal distances.

(*b*) Miller considered Tubar as an isolated language and asserted that "Cáhitan forms a link between Tubar and Corachol" (1984:17). It would be difficult to link Tubar with Taracahitan despite its greater resemblance to Cahitan than to Tarahumaran or Opatan. Although we generally support Miller's proposal, we do suggest one argument against his assumption concerning Tubar: the best dendrogram (the one with the greatest internal homogeneity, obtained by SSC) shows Tubar to be coordinated with Cahitan rather than with the Taracahitan group. Nevertheless, the MDS's always mapped Tubar as closer to Cahitan but not sufficiently so to enhance the relationship. Also, in some K-MEANS clusterings, Tubar was linked with Cora, thus suggesting that its distance from the core Sonoran is sufficiently long to dilute its relationship with Taracahitan.

Aztecan. The graphs clearly show two branches within Aztecan: Pochutec and the rest of Aztecan, which could in turn be further subdivided.[14] This clustering shows clearly in all our graphs. We can assume that Pochutec is a distinct branch on the basis of the longer internal distances that exist within Aztecan. If we compare those distances (for instance, in figure 2), we must say either that Aztecan is not a tight family or that Aztecan has at least two groups. We prefer the second option.

(*C*) The tree representation proved to be not at all applicable to the Uto-Aztecan classification. The hierarchical structure suggested by the dendrograms does not adequately represent the real relationships involved in the distance matrix. Although at some certain levels of representation, the tree can be used as an almost satisfactory model, it cannot

[14] Lastra (1986) has made an excellent attempt to classify the internal differentiation of Aztecan, although we regard this thus far as preliminary.

be used as an adequate representation of the family as a whole. Additional data analysis methods should be used, as Miller has suggested.

It is clear that a more refined Uto-Aztecan classification must involve many other variables and additional data. Insofar as we try to depart from strictly lexical evidence, we should have more data, ideally with at least three samples of each language. Even so, it is unlikely that the results would be adequately represented by means of a tree; but the lexical data could give more information than we have predicted and could provide a basis for deciding among various alternative potential classifications.

Finally, it is important to stress that the statistical methods are not an end by themselves; they allow us to gain a better understanding of the interrelations that otherwise are difficult to find. We need to know how to evaluate their advantages and disadvantages in order to benefit from them.

APPENDIX A

Cognate Density Matrix of the Six Languages Added

	Mo	PN	Pa	Sh	Cm	Ka	Ch	PS	Ut	Tb	Se	Ca
Southern Tepehuan	22	24	23	23	21	24	24	26	27	33	28	31
Tepecano	13	15	13	15	13	16	17	20	20	33	30	28
Mexicanero	16	15	12	12	11	11	11	12	11	20	14	18
Tuxpan	14	14	13	13	12	15	13	15	15	19	15	18
Michoacán	18	17	14	14	13	14	14	15	14	21	16	18
Classical Nahuatl	18	18	16	16	14	16	15	16	16	23	22	24
Telecingo	18	18	16	16	14	16	15	16	16	23	22	24
Zacapoaxtla	17	17	15	15	13	16	16	17	17	24	20	24
Pochutec	12	12	11	10	8	12	16	12	11	17	14	18
Pipil	16	16	14	14	12	6	15	16	16	23	19	21

	Cu	Lu	Ga	Ho	Pg	Ne	TN	Op	Eu	Gu	Ta	Tu
Southern Tepehuan	28	25	26	30	73	75	82	39	43	48	43	43
Tepecano	24	22	26	24	76	78	87	45	48	54	46	48
Mexicanero	17	17	15	18	21	20	22	25	28	23	24	23
Tuxpan	16	16	14	20	27	27	29	28	31	25	26	32
Michoacán	16	16	15	20	24	23	26	28	31	25	26	29
Classical Nahuatl	19	20	19	24	29	29	30	36	41	33	33	36
Tetelcingo	19	20	19	24	29	29	30	35	39	31	32	34
Zacapoaxtla	19	20	20	25	31	31	32	33	38	29	32	35
Pochutec	16	14	15	19	23	22	25	28	28	24	23	25
Pipil	18	18	19	25	29	29	30	33	36	31	31	35

	My	Yq	Co	Hu	TS	Tc	Mx	Tx	Mi	Nc	Te	Za	Po
Southern Tepehuan	49	49	35	41									
Tepecano	54	52	37	49	89								
Mexicanero	26	27	28	27	22	24							
Tuxpan	31	32	29	30	27	29	57						
Michoacán	31	32	30	30	23	24	67	65					
Classical Nahuatl	39	40	37	39	29	33	66	65	76				
Telecingo	37	38	34	36	29	33	64	66	73	85			
Zacapoaxtla	38	39	33	35	31	37	57	64	66	78	81		
Pochutec	31	32	25	27	24	28	45	54	53	54	55	55	
Pipil	40	41	33	35	29	33	59	66	68	78	78	75	57

REFERENCES

ANDERBERG, MICHAEL. 1973. Cluster Analysis for Applications. New York: Academic Press.

ARDISSON, MÓNICA. 1980. El programa Hartigan: manejo de un paquete de técnicas de clasificación. Manual de usuario. Mexico: IIMAS-UNAM.

BOAS, FRANZ. 1917. El Dialecto mexicano de Pochutla, Oaxaca. IJAL 1:9–44.

D'AGOSTINO, RALPH, AND E. S. PEARSON. 1973. Tests for departure from normality: empirical results for the distributions of b2 and $\sqrt{b1}$. Biometrika 60, no. 3:613–22.

EDWARDS, A. W. F., AND L. L. CAVALLI-SFORZA. 1965. A method for cluster analysis. Biometrics 21:362–75.

ESPINOSA, GUILLERMO, AND ARTURO LÓPEZ. 1977. Introducción a los métodos jerárquicos de análisis de cúmulos. Mexico: Comunicaciones Técnicas, IIMAS-UNAM.

ESPINOSA, GUILLERMO; LILIA REYES; AND ARTURO LÓPEZ. 1978. ANALISIS/CUMULOS: un programa para análisis de cúmulos. Mexico: Comunicaciones Técnicas, IIMAS-UNAM.

GORDON, A. D. 1981. Classification. London: Chapman and Hall.

HARTIGAN, JOHN. 1975. Clustering Algorithms. New York: Wiley.

IRAIZOS, MANUEL. 1984. Implantación de programas software para realizar técnicas de escalamiento multidimensional. B.S. thesis, Faculty of Sciences, UNAM, Mexico.

KEY, HAROLD. n.d. Vocabularies for languages of the Uto-Aztecan family: introduction to Aztec-Tanoan. Microfilm collection 7th, no. 38.

LASTRA, YOLANDA. 1986. Las áreas dialectales del náhuatl moderno. Mexico: IIA-UNAM.

LUMHOLTZ, CARL. 1904. El México desconocido. México: Instituto Nacional Indigenista.

MARTÍNEZ MALO, LUZ MARÍA. 1979. Algunos métodos jerárquicos y otros subdominantes de taxonomía numérica. B.S. thesis, Faculty of Sciences, UNAM, Mexico.

MASON, J. ALDEN. 1917. Tepecano: a Piman language of Western Mexico. Annals of the New York Academy of Sciences 25:309–416.

MILLER, WICK R. 1984. The classification of the Uto-Aztecan languages based on lexical evidence. IJAL 50:1–24. [First presented at the Symposium on Uto-Aztecan Historical Linguistics, University of New Mexico, June 1980.]

SÁNCHEZ-OLMEDO, JOSÉ G. 1980. Etnografia de la Sierra Madre Occidental: Tepehuanos y Mexicaneros. Mexico: Instituto Nacional de Antropologia e Historia.

SHEPARD, R. N.; A. K. ROMNEY; AND S. B. NERLOVE. 1972. Multidimensional Scaling: Theory and Practice in the Behavioral Sciences. New York: Seminar Press.

SWADESH, MORRIS. 1955. Towards greater accuracy in lexicostatistic dating. IJAL 21:121–27.

VALIÑAS, LEOPOLDO. 1981*a*. El náhuatl de la Periferia Occidental y de la Costa del Pacifico. B.A. thesis, National School of Anthropology and History, Mexico.

_____. 1981*b*. Los préstamos lingüísticos como diagnósticos regionales. Los corporales en Durango y Nayarit como ejemplo. Anales de Antropología 18, no. 2:227–37. Mexico: UNAM.

VALIÑAS, LEOPOLDO, AND MARIO CORTINA-BORJA. Forthcoming. Contribución a la reconstrucción histórica a partir del análisis estadístico en datos léxicos: el caso de las lenguas sonorenses. Memorias del XII Congreso de Antropología e Historia de Sonora, Hermosillo, Sonora, México.

_____. In preparation. An attempt of historical reconstruction from lexical evidence: the Southern Uto-Aztecan languages.