Ever since the initial outlining of the HOKAN and PENUTIAN stocks (Dixon and Kreber, 1913), the two, although geographically intermeshed, have been classified apart, cut to the level of such wide-ranging speculative groupings as Sapir’s (1929) six North American superstocks, in HOKAN-SIOUAN and a much expanded PENUTIAN. Only when the speculation ranges even farther out to networks encompassing almost all the languages of North America are the two linked (as in the work of Swadesh, for example, 1959). Yukian, too, has been kept distinct at the stock level; but it has been joined to others at the more remote superstock level: to HOKAN, on the one hand (for example in HOKAN-SIOUAN, Sapir, 1929); to PENUTIAN, on the other (as suggested in Radin, 1919, and Shipley, 1957). At this point, I should inject a word of caution for those who might be unfamiliar with the type of evidence that supports the postulated groupings HOKAN, PENUTIAN, and YUKIAN (to say nothing of the more far-ranging proposals)—it is skimpy. The divergence between the branches of HOKAN, and of PENUTIAN, is much greater than that typical of the branches of Indo-European. The two branches of YUKIAN—WAPPO in the south, and the YUKI-HUCHNOM-COAST YUKI dialect cluster in the north—are considerably more alike lexically (about like FINNISH and HUNGARIAN), but still not so close as to be incontestably related (Sawyer, in press, argues against a genetic relationship).

However, it is not my intention here to review the proposals of others but rather to offer an exploration of the affinity among representatives of the three stocks, employing explicit procedures for measuring lexical similarity.

The deepest exploration is among the following four:

WAPPO (YUKIAN). The first 100-word list is from Sawyer (1962), the second 100-word list from Sawyer (1965), with segmentation, mostly to CVC roots, helped by personal consultation with Sawyer.

PROTO-WESTERN POMO (= PWP, HOKAN). The reconstructions are based on my field notes in the four most closely related POMO languages: Kashaya, Central, Southern, and Northern POMO. Many of the reconstructions could also be projected back, with no change, to Proto-POMO.
Proto-Western Miwok (=PWM, Penutian). The forms were compiled from Callaghan (1965, 1970, and personal communication). Many are also good for Proto-Miwok.

Maidu (Penutian). The forms are principally from Shipley (1963), but some attempt was made to use Proto-Maidu by taking into account the forms in the closely related Nisenan (Uldall and Shipley, 1966).

For each of the above four linguistic entities, two distinct lists of 100-words were compiled, in order to test the replicability of the results from one sample of vocabulary to another. Both word lists were relatively "basic" because I wanted any affinity discovered to be more likely to reflect a genetic relationship than a recently diffused one. The first 100-word list is the Swadesh short list (given, for example in Hymes, 1960) and is relatively universal and culture-free. The second 100-word list is my own selection and contains kinship terms, body part terms (additional to those in the first list), common adjectives and verbs, and some terms for flora and fauna (corn, deer, etc.) which, although not of world-wide distribution, are common and basic in the area under investigation. To a certain extent I was guided by what I could reconstruct.

In addition to the results with the above entities, there will be presented findings from earlier investigations, based on the Swadesh 100-word list alone, for comparisons of Wappo with Tonkawa, a "Hokan-Coahuiltecan" language of Texas (data from Hoijer, 1949), and the four Hokan languages Washo (data from William Jacobsen, Jr., personal communication), Karok (from Bright, 1957), Cocopa (from James Crawford, personal communication), and Kashaya, a language also involved in the Western Pomo reconstructions. The findings on the interrelationships of these languages which do not concern Wappo will appear in a separate publication (and were also discussed in Oswald, 1965, 1966).

As stated earlier, any affinity among Wappo, Hokan, and Penutian (and even within Hokan and Penutian) would be at a very distant level, beyond that of conventional proof: any possible cognates would be too few in number, and the sound changes too great, to permit setting up a comprehensive interlocking scheme of sound correspondences. In such situations, there have been, in North America and the rest of the world, speculations on relationships among diverse languages, with the "proof" being a listing of vocabulary similarities. However, such similarities can be found between any two languages, related or not, and the number may depend less on the degree of relationship than on the volume of material available to the investigator and on what he is willing to call sufficiently similar, semantically and phonetically; factors which are usually left unexpressed and variable.
The method employed here goes beyond a mere search and listing of vocabulary resemblances: The criteria of similarity are standardized and explicit. The number of similarities thus found are then tested for significance by being projected against a background of the number and range of similarities that can be found by chance.

The criterion of semantic similarity is simple: the word employed should be the most common one for the meaning given. There are occasional difficulties in the selection of the proper word, especially when a dictionary must be used rather than a native speaker or a linguistic expert in the language (dictionaries rarely state which of several words translated the same is the most common and general), but in these particular cases no crucial problems arose. All forms were segmented as much as possible, consistent with retaining the desired meaning. I would now, in principle, eliminate from any particular comparison those forms which, in one language or the other, require more than one morpheme to attain the desired meaning, but that procedure was not followed here. Thus I used a few bimorphemic forms like Maidu wóno-ti 'die-cause' for the Swadesh gloss 'kill'.

The determination of phonetic similarity is more complex:

1. In order for paired forms to be counted phonetically similar, a threshold number of consonants and vowels must match. In all the cases presented herein, the number of required consonant matches is two if the forms in each language have two or more consonants. If the form in one or the other language has only one consonant, then only one consonant match is required. There are no cases of no consonants. In the comparisons in this paper, vowels are not considered; this is not for theoretical reasons (although I do believe that vowels are of less utility than consonants in judging cognacy), but because of temporary technical factors (discussed in Oswalt, 1966, 1970).

2. For two consonants to match, they must share
   A. a point of articulation.
   B. a threshold number of articulatory features.

The point of articulation (Criterion A) can be interpreted broadly or narrowly: Since some of the languages dealt with herein make a phonemic differentiation of alveolars from dentals, but most do not, I usually interpreted these points broadly and equated the two—and similarly for the velars and uvulars. However, in the case of PWP and Wappo, the languages both contrast voiceless stops (but not n and d) at the dental and alveolar positions; I treated these points narrowly, allowing matches, for voiceless stops, of only dental to dental and alveolar to alveolar.
Both the articulatory features to be considered and the threshold number were varied in different trial comparisons. A typical criterion b is that the two consonants share two of the three features of voicing, nasality, and friction (this last is a plus feature for fricatives and aspirated stops). The consequences of these simply stated criteria are illustrated below; for each of PWP labials listed at the left, there are given at the right all the Wappo consonants that would be accepted as matches:

\[
\begin{array}{l}
b & p, \, p', \, m, \, \ddot{m}, \, w, \, \ddot{w} \\
p & p, \, ph, \, p' \\
ph & p, \, ph, \, p', \, w, \, \ddot{w} \\
p' & p, \, ph, \, p \\
m & m, \, \ddot{m}
\end{array}
\]

There were various niceties of application which will not be gone into; one example applies in PWP: h and \(\ddot{o}\) before a vowel are counted as consonants (laryngeals), but before a consonant (as in \(\ddot{b}i\ddot{v}du\ 'acorn'\) are nonconsonants.

When the first 100-word lists in PWP and Wappo are lined up, semantically matched, 12 pairs of words are found to pass the above stated criteria for phonetic similarity. The question then is, "Is 12 out of 100 sufficient to claim a special affinity between the two languages?" In order to determine this, a SHIFT TEST can be performed; that is, the first word in PWP can be compared to the second in Wappo, and the second in PWP to the third in Wappo, and so on down the list until the 100th word in PWP is brought around to be compared to the first in Wappo. When this is done, under exactly the same criteria of similarity, a score of 7 is obtained. Since the words are no longer semantically aligned, this score is one measure of the number of chance resemblances that can be expected between the two languages, A BACKGROUND SCORE. Another question is, "How reproducible is this background score?" The test can be performed again by shifting the lists two places relative to each other, and a third time by shifting three places, and on up to a maximum of one less than the length of the lists (99 in this case). Such a procedure is quite time-consuming, mind-numbing, and fraught with human error, when done by "eyeballing"; a rather complex computer program was developed to perform the job and used for the findings reported herein.
The results from a comparison of Proto-Western Fomo and Wappo, under the criteria described, for 99 shift tests are plotted in Figure 1. Each x represents one shift test. The position of an x on the horizontal axis indicates its score; the height of a stack of x's gives its frequency, the number of times that the particular background score occurs. For example, the score 2 occurs 2 times; the score 3 occurs 12 times. The BACKGROUND MEAN, the average of all the background scores, is 6.06, located by the vertical arrow near the high point of the plot. The GROSS SCORE, the score obtained when the lists are compared semantically matched, is 12, marked by the vertical arrow toward the right of the plot. One estimate of the significance of a score of 12 is given by its RANK: Of the 100 scores, 12 lies in third place; two background scores are higher, 97 are lower. This is moderately strong evidence that there is some kind of special affinity between the two vocabularies but that some caution is called for—illustrated graphically with two chance background scores higher than the gross score. More refined ways of estimating the significance will be given later, but first there will be plotted the results with this same pair of languages and vocabularies, under different criteria of similarity.
In Figures 2 and 3, Wappo and PWP are compared under the same conditions as before, except for Criterion 2B: in Figure 2, the consonants must match in 3 of the 5 features: voicing, stoppage, nasality, friction, and laterality; in Figure 3, they must match under the stricter criterion of 4 of those same 5. Since only one of the many different consonants in these languages is plus on laterality (namely, l), this feature is infrequently brought into play and, in most instances, the requirements are to match 2 of the first 4 features, and 3 of 4, respectively.

Under the less stringent criterion of Figure 2, the gross score of 14 ranks highest, above all 99 background scores; this is even more powerful evidence of a special affinity between the two languages than was displayed in Figure 1. In Figure 3, under much stricter conditions for passing the threshold of similarity, the background scores are lowered, as should be expected, but the gross score is lowered relatively more, so that it does not stand clearly above the majority of the background scores—and this is the situation that often prevails among languages which are, at best, only remotely related.

The difference between the gross score and the background mean is the GROSS DEVIATION; in Figure 2, it is 14 - 6.2 = 7.8. Certain adjustments can be calculated for the gross deviation to yield the NET SCORE, an estimate of the number of words in the list whose similarity cannot be attributed to chance and for which an historical explanation should be provided (the calculation is explained in Osvartz, 1966, and will not be described here, as the discussion herein will be based on other measures of affinity).

In Figures 1-3, the x's fall into a normal distribution about as well as most natural data. In Figure 4, the plots are converted into a standard normal curve in which the horizontal axis no longer represents numbers but STANDARD SCORES, with which the unit of measurement is the STANDARD DEVIATION (the square root of the mean of the squares of the deviation of each background score from the background mean).
The standard deviation in Figure 1 is 2.29, in Figure 2 it is 2.35, and in Figure 3 is 1.65. The gross deviations in the three cases can be converted to standard scores by dividing by the standard deviations of their respective comparisons:

1. \[(12 - 6.03)/2.29 = 2.61\]
2. \[(14 - 6.22)/2.35 = 3.31\]
3. \[(5 - 2.60)/1.65 = 1.45\]

These three scores can now be located on one graph, Figure 4, in comparable terms. The background means, lying at the peak of the curves, are set at zero, regardless of what they were in Figures 1-3; the standard units along the horizontal axis correspond to different word scores in each of the three cases.

For each location along a normal curve there is an associated probability represented by the ratio of the area under the curve that lies to the right of the given location to the total area (the information is available in many statistics books in tables of areas under the standard normal curve). For Case 1, the probability of getting a standard score of 2.61 or higher, by chance, is 0.0045, that is, four and a half in a thousand. For Case 2, the probability of a score of 3.31 or higher, by chance, is 0.0005, only five in ten thousand. The minuteness of these probabilities is graphically illustrated in Figure 4: the points 2.61 and 3.31 lie to the right of almost the entire area of the curve. These probability measures are so low that it would be wise to discard any hypothesis of chance as the explanation for the resemblances found between Wappo and PWP and propose some kind of historical relationship, either genetic or diffused. This proposition will be explored later in this paper, but first I should give a few words on the effect of varying the criteria of similarity.

A great deal of effort has been expended on an investigation (which will not be detailed here) of the effect of varying the criteria of similarity; one example is the series 1 to 3 above. This effort was partly to test an hypothesis that the procedure would help find the basic sound correspondences among distantly related languages, correspondences which are not obvious because possible cognates are so few and are hidden in a mass of chance resemblances. I cannot now say that the procedure will prove to be feasible, although it does work between languages that are closely related--there than is peak in the standard score (= a minimum in the probability measure) when the criteria of similarity closely approximate the true sound correspondences.

There is a difference between closely related languages, vs others, in response to strictness of the criteria of similarity. With closely related languages, the sound changes are relatively slight and a high threshold (such as requiring matching of 4 of 5 features, rather than 3 of 5) causes
rejection of false matchings to a greater degree than it causes failure to detect true cognates; that is, the background is lowered more than the gross score, and thus the gross deviation and its associated standard score are raised (this is illustrated for English-German and Persian-Hindi pairings in Oswalt, 1970). With more distantly related languages, the effect can be the opposite; the sound changes may be great enough that a high threshold will cause the rejection of true cognates as much as, or more than, it lowers the background. And the latter pattern is apparently the one that the Wappo-PWP comparison falls into (judgment is withheld for now on whether the large number of resemblant forms are true cognates or diffused forms); the high threshold case (3) has a markedly lower standard score than the low threshold case (2). Case number 1 is intermediate in threshold and standard score, but it focuses on a smaller number of features, three (voicing, nasality, and friction in this case; different selections of three in other cases not presented here), and thus is more subject to the idiosyncratic characteristics of particular word lists.

Now that the method has been described in a limited, but what I hope is sufficient, way, the results of the remainder of the comparisons can be presented in a compact form. In Figures 5 and 6, the graphs are essentially the same as in Figure 4, the difference being that the normal curve is omitted and the bottom line alone retained, a line in which the zero in the middle is the background reference point and the standard scores (of the gross deviations) for each comparison are placed in their proper location.

Figure 5 contains the comparisons for the first 100-word list (the Swadesh list) for all possible pairings of Wappo, Proto-Western Pomo, Proto-Western Miwok, and Maidu, plus pairings of Wappo with Washo, Karok, Kashaya, Cocopa, and Tonkawa. The criteria of phonetic similarity are the same as for Figure 2, the least stringent of those exemplified earlier.

An easy decision that one can make on the basis of Figure 5 is that there is no discernible affinity between any pair of languages that displays a negative score (occurring when the gross score is less than the background mean). It is harder to decide how large a positive score should be before it deserves to be deemed significant. A commonly employed significance level is 5 percent, that is, the investigator is willing to be wrong 5 times out of 100 in claiming the results are not due to chance. This level corresponds to a standard score of 1.645. I consider it somewhat low for proposing linguistic relationships and in this opinion am influenced by the -1.66 in Figure 6 (and by similarly large negative scores obtained in comparisons not introduced here). Since, in the 17 comparisons displayed here, -1.66 is reached by chance (and I cannot imagine a reasonable historical explanation for a negative relationship), I would expect +1.66 to be as easily attained by chance. I am still acquiring experience toward a good, working significance level for
Figure 5. First 100-Word List

Figure 6. Second 100-Word List
linguistic comparisons and at the present take a conservative position that the standard score should be greater than 2.00 before suggesting a relationship and above 2.50 before making a strong claim. In this paper, the point is moot as there are no scores in Figures 5 and 6 between 1.645 and 2.00.

In Figure 5, in the suggestive range 2.00 - 2.50, there are three pairings of Wappo: one with the neighboring PWM and two with distant Hokan languages, Washo and Cocopa. Lying far above these is the strong score with the neighboring PWM and its member Kashaya, not geographically adjacent to the Wappo.

In order to test these findings, a second set of 100-word lists was prepared and run under the same criteria of similarity. The results were different; no pairs of word lists were found to be significantly alike (Figure 6). This is perhaps not so surprising in the less strong case Wappo-PWM (Kash) but it is striking for the Wappo-PWM case. The Wappo-PWM and Wappo-Kashaya scores with the first 100 words are so strong that they require some historical explanation regardless of the lack of affinity evinced with the second 100 words. The scores of 3.31 and 3.48 can perhaps be given some perspective by a comparison: they are higher than any score obtained between branches of Hekan (Oswalt, 1966) and are in the range of scores between modern representatives of separate branches of Indo-European, such as German and Russian (Oswalt, 1976).

The two word lists are not randomly drawn from a large vocabulary pool, nor are they even a random splitting of a 200-word list—if either had been the selection procedure, the discrepancy between the two results for Wappo and PWM would have been more extraordinary. The two vocabularies were chosen for "basicness" in order to minimize susceptibility to borrowing, but they are not equally basic, and therein may lie the explanation.

I have long had the goal of being able to arrange a large semantic list, say of 500 items, in order of relative stability. If, when batches of this graded vocabulary are taken in sequence for comparison, a trend should show up such that the relatively more stable batches exhibit significantly more affinity than the less stable batches, then one might infer that the paired languages are genetically related; if the trend runs in the opposite direction, one might infer that the affinity has resulted from borrowing. This hypothesis would, of course, have to be tested with languages known to be related and with languages known to be not genetically related but to be influenced by one another. A start in measuring the relative stability of various semantic items was made in Oswalt (1971, 1975) but the project is so immense that it cannot be expected to progress far in one lifetime.

Of the two vocabularies employed here, the Swadesh list must be considered the more basic and stable overall (but not so for certain items); it, after all, was first choice, and the second list contains leftovers. Now, the Wappo-PWM comparisons follow the pattern that the supposedly more basic list shows the stronger score and the surmise could thus be that the
affinity between the two is genetic. However, I withhold any such strong claim because the results with the two word lists are so disparate, two batches seem too few to establish a trend, and the hypothesis is insufficiently tested as yet. It would also be wise to bring Yuki into a similar comparison (See Note 1), for a claim to be extendable from Wappo to Yukian.

Since PWP and PWM are geographically close to Wappo (although Kashaya is not in direct contact with Wappo), it would be reasonable to suspect a certain amount of diffusion among them. Perhaps an examination of the actual resemblant forms will lead to an understanding of the source of the high scores. Below are given the sets of forms from the first 100-word list that figure most prominently in the resemblances between Wappo and PWP (plus Kashaya), and in the other pairs that score over +2.00 (PWM, Washo, and Cocopa) if they fit into these sets, otherwise forms from these other languages will not be cited beyond the first four sets.

( . . . ) encloses forms found similar to some of the others in the set, but not to the Wappo form.

[ . . . ] encloses forms that are not similar to any others of the set.

# marks reconstructions to the level named at the left: PWP or PWM.

** marks reconstructions good to a deeper level: Proto-Pomo or Proto-Miwok.

<table>
<thead>
<tr>
<th>BIRD</th>
<th>EAR</th>
<th>THOU</th>
<th>ROUND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo</td>
<td>čī ča</td>
<td>čéma</td>
<td>mí*</td>
</tr>
<tr>
<td>PWP</td>
<td><strong>šihta</strong></td>
<td><strong>šima</strong></td>
<td>(<strong>šama</strong>)</td>
</tr>
<tr>
<td>Kashaya</td>
<td>šihta</td>
<td>šima</td>
<td>ma</td>
</tr>
<tr>
<td>PWM</td>
<td>[<em>méle</em>]</td>
<td>[<em>šišak</em>]</td>
<td>*ši-</td>
</tr>
<tr>
<td>Washo</td>
<td>sī-su</td>
<td>ŕisew</td>
<td>ma</td>
</tr>
<tr>
<td>Cocopa</td>
<td>šá</td>
<td>š mál</td>
<td>(mapú)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DOG</th>
<th>SKIN</th>
<th>SUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo</td>
<td>háyu</td>
<td>chíla</td>
</tr>
<tr>
<td>PWP</td>
<td>? háyu</td>
<td><strong>ši'ıda</strong></td>
</tr>
<tr>
<td>Kashaya</td>
<td>háyu</td>
<td>ši'da</td>
</tr>
<tr>
<td>PWM</td>
<td><strong>šá háyu</strong></td>
<td><strong>šú luk</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DRINK</th>
<th>I</th>
<th>LONG</th>
<th>WOMAN</th>
<th>DRY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo</td>
<td>?ú</td>
<td>?áh</td>
<td>fína</td>
<td>mé te</td>
</tr>
<tr>
<td>PWP</td>
<td><strong>hòqo</strong></td>
<td>*<em>há'a</em></td>
<td><strong>šiçöl</strong></td>
<td><strong>šima'ta</strong></td>
</tr>
<tr>
<td>Kashaya</td>
<td>?qo</td>
<td>?a*</td>
<td>?ahqol</td>
<td>?ima' ta</td>
</tr>
</tbody>
</table>

66
BIRD. The similarity of these forms could be due, in part, to onomatopoeia, by a mechanism analogous to an English speaker calling a small bird a tweet, peep, cheep, chirp, etc.

EAR. This is one of the very few sets of resemblant forms widespread in Hokan. See below for a comment on glottalization.

THOU. m- in the second person pronoun is widespread in Hokan and Penutian. In this set, the parenthesized forms do not match with the Wappo because the two-consonant criterion applied. I would now, on the basis of information internal to the languages, segment the two-consonant forms down to the one consonant, plus the vowel. This action would increase the scores among the three affected languages over that reported here. In the Proto-Pomo form, for example, the initial syllable *φa- can be considered a separate morpheme marking pronouns and deictic forms. Compare also Proto-Pomo *ha- in the 'I' set.

ROUND (spherical rather than circular, if a particular language makes a distinction). Similar forms are widespread in central California, and no one language or family can be said to be the source or the recipient of a borrowing for the term. Since a word ROUND cannot be reconstructed for PWP, one might suppose it to be a loan into Kashaya, but I think not; it can be explained as arising from a vast complex of symbolic forms (many of which occur in other Pomo languages): lo'lo' 'cylinder', ko'lo' 'concave', tōlo'lo' 'dimple', dolo' 'mortar basket', dolō- 'to make a depression', polo- 'oak ball', polopolo- 'to bubble', and a host of verb roots like -bo- -bo' --bolh- -bo'm- -bo' -ho- -po- -po' -etc., all having something to do with swelling up and becoming spherical, or being hollowed out and becoming round on the inside. Furthermore, there are the roots -lo- 'to move spherically, to roll (of a ball)' and -li- 'to move circularly'. And the root -li- leads us to pili'li 'circle, hoop', the analog of our starting point polo'lo'.

Insofar as the dictionaries and vocabularies of the languages of central California go into sufficient detail, they also attest to similar symbolism (and the symbolism goes pretty far in the rest of the world too): Wappo polo'k 'oak ball'. Lake Miwok póllolo, pówwolo, pówwu' 'round'; pololo' 'ball, dry oak ball'. Maidu pyłylm 'round', pyłylmpe 'spherical', pólpó 'bubble, boil, steam'. Nisenan pylyl 'spherical', polo' 'buckeye ball', molmol 'to boil'. Washo múlmol 'to boil'.

Much further to the south, in Cocopa and many Yuman languages, the forms for ROUND are merely tantalizingly close. Taking C1VC0 as the canon of the core of the symbolic form (pol in central California), we find in Yuman that C0 is coronal, varying symbolically for size and intensity (Langdon, 1971). Since C0 is often a lateral, it would match the C0 of pol; however, the C1's are usually different and thus the total forms fail to match: Yuma *ora^or to be spherical'; Tipai
yaryar, yai¥ya¥ 'to be round'. Closer phonetically is a Diegueño complex for 'roll': mul, mil, mili, mil (I have extracted what I judge to be the core, I hope correctly, from entries in Cocurc and Hutcheson, 1973); mul¥mul wi: 'is round like a ball'.

It should be noted that the Wappo form álbul failed to match the others in the ROUND set, although it fulfills the criteria of phonetic similarity, and this is due to BLOCKAGE. The consonants of the Washo form are taken in sequence to compare with the consonants of the other form (Wappo -pól-, for example) also considered in sequence. The first l in álbul is found similar to the l of the Wappo form; then the b in the Washo must find its match after that matching pair, and there is none in -pól-. If the first l in álbul were then eliminated from consideration, the -bul sequence of Washo would fit very well with the Wappo -pól-. Consonant blockage causes the rejection, on average, of about one and a half suitable matches in a 100-word list. This lowers both the gross and background scores about the same so that the net effect is small, but it is a factor in the reduction of the sensitivity of the method (blockage was eliminated from the program but could not be applied to these data because of computer use).

DOG. This is a widely diffused form in the region; I am not sure of the source, possibly Penutian (compare Tolitich Yokuts xalu 'coyote', cited in Shipley, 1957:271). It probably did not exist in Proto-Pomo (there are aberrations in the sound correspondences); nevertheless I did use hayu in PWP comparisons. Omitting the word from the vocabulary would lower the scores slightly.

SKIN. I would now segment the final -k of PW¥gúluk, as well as Proto-Miwok gálok 'ear', as a noun-formative. Doing so would not change the gross scores but would lower the background means and thus raise the gross deviations.

DRINK. There is a fair chance that this originated as an imitative of the action and sound of noisily swallowing a liquid, that is, of gulping. In Kashaya, ḫו- 'to drink' is connected in a symbolic and echoic set with ḫu 'sound of a gulp' (and also the root in a verb ḫu 'to gulp'), ḫol 'sound of gurgling (also in the verb ḫolo- 'to gurgle'), and ḫol 'sound of lapping water'. It may well be no accident that English gulp, gurgle, gargle, etc., are similar to these Kashaya forms. In any case ḫo- can be reconstructed for Proto-Pomo with perfect sound correspondences.

Historical reasons, other than genetic relationship, have now been suggested for the resemblances in some of the above sets, namely: onomatopoeia with strong elements of universality (BIRD, DRINK), a diffused symbolic system with a slight element of universality (ROUND), and plain borrowing (DOG). As evidence for a genetic relationship, these sets are thereby weakened, but they cannot be totally ignored--these possible, even probable, alternative origins are not certainties.
Even if these four sets be eliminated from consideration, there still remain resemblant sets in somewhat greater number than expected by chance. A goal might be to separate out the stronger candidates for cognacy. As one attempt, it might be argued that those sets which rely on glottalized consonants with unglottalized in an unsystematic way are vulnerable. However, we are dealing here with a tremendous time-depth, during which many changes have taken place—and glottalization can arise other than through some obvious phonological development; it might result from morphological affixation and a blending of consonants (this has occurred with some stem-final consonants in Kashaya); it might result from a symbolic alteration in a sound. I might add that in a more "classical proof" of cognacy it is quite easy to find recurrent sound correspondences between Wappo and PWP (or Kashaya):

<table>
<thead>
<tr>
<th>EAR</th>
<th>THOU</th>
<th>WOMAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo m</td>
<td>déma</td>
<td>mi-?</td>
</tr>
<tr>
<td>Proto-Pomo m</td>
<td>***shima</td>
<td>***a-ma</td>
</tr>
</tbody>
</table>

Wappo has no voiced stops (except in Spanish loans) and it is thus also reasonable to postulate that PP **b (or Kashaya b) corresponds to Wappo m. For examples of this it is necessary to go to the second 100-word list:

<table>
<thead>
<tr>
<th>ACORN</th>
<th>SHORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo m</td>
<td>mé l</td>
</tr>
<tr>
<td>PWP **b</td>
<td>**bi-du</td>
</tr>
<tr>
<td>Kashaya b</td>
<td>bi-du</td>
</tr>
</tbody>
</table>

In parallel fashion one could postulate that PP**d : Wappo n, but for this there is in these lists only one example SUN PP **na#da : Wappo hin; while for an equally reasonable correspondence of PP **d : Wappo l, there are two good examples, the above cited ACORN and SKIN PP **di#da : Wappo chila.

If we allow the correspondence to go slightly beyond the criteria of similarity employed herein, we could have

<table>
<thead>
<tr>
<th>EAT</th>
<th>FATHER'S MOTHER</th>
<th>FATHER'S SISTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wappo p</td>
<td>pa?-</td>
<td>-pápa</td>
</tr>
<tr>
<td>Proto-Pomo **m</td>
<td>**ma?a-</td>
<td>**-ma-</td>
</tr>
</tbody>
</table>

More examples could be brought forth and many could be made to fit into some sort of system by ad hoc, environmentally conditioned rules or by devising new proto-symbols. However, given the great divergence of the languages treated herein and the masking effect of the chance resemblances, which are about as numerous, in the Wappo-PWP comparison, as the expected number of forms historically related from all causes, it is doubtful that it will ever be possible to assemble a
reasonable, convincing, encompassing, multiply-attested system of reconstruction.

The approach in this paper is quite different; it does not depend upon being able to separate potential cognates from chance similarities and borrowings and so to reconstruct. It depends on a statistical analysis of semantically paired vocabularies in order to test whether the phonetic similarities found between them are significantly greater than can be attributed to chance. The statistical analysis cannot of itself distinguish an affinity arising from a common genetic origin from an affinity resulting from other historical causes, but proof of an historical connection of any type would be a worthwhile advance in our knowledge of the interrelationships of American Indian languages. With the pair showing the strongest affinity, Wappo-PWP, an attempt was made, through an investigation of the individual sets of semantically and phonetically matched forms, at a partial separation of the resemblances due to different historical causes.

No significant affinity between PWM and Maidu shows up in these comparisons, a reflection of the problematic status of Penutian. There is a tilt of Wappo towards the languages called Hokan, especially the Pomo family, more so than towards the two languages called Penutian; but a strong claim of a genetic relationship between Wappo and the Pomo family is held in abeyance pending a deeper exploration with more vocabulary. Any more comprehensive claim that the grouping known as Yukian is related to another stock, family, or language must rest on comparisons that also include a vocabulary representative of the Yuki-Huchnom-Coast Yuki dialect cluster, as well as the vocabulary of Wappo.
FOOTNOTE

1. This paper was presented orally at the 1977 Hokan-Yuman Languages Workshop held at the University of Utah. The comparisons involving PWP, FWM, and Maidu were done in 1966; the others date from 1963-1965. Both sets of work were done through the facilities of the Computer Center of the University of California, Berkeley.

Occasional reference is made throughout the paper to courses of action that would clarify the results or improve the procedure technically, sometimes with a referral to this note, sometimes not. Sample suggestions are:

Better segmentation of the forms.
More batches of vocabulary.
Inclusion of more putatively related languages: Hokan, Penutian, and especially Yukian.
More detailed investigations of the effects of varying the criteria of phonetic similarity.
Refinements in the machine program.

A lot of additional data have been assembled and processed to run, including word lists in Yuki, but all work with the computer came to an abrupt halt in 1968, when the Computer Center replaced its earlier generation machine with a more advanced model. The program was developed over a period of years, in assembly language, to run on the IBM 7090 and will not work on the later machines. It will be a long, expensive, complex task to rewrite the program and, for the past ten years, I have not had sufficient access to a computer powerful enough to handle the projected program. With the great improvements in microcomputers, and their dramatic drop in price, the situation may change in the near future. Pending the resumption of that line of endeavor, I must present the results "as is" regardless of the ways I see to make technical improvements. Others are welcome to try their hands at the method.
BIBLIOGRAPHY

Bright, William, A Reverse Index of Yuma Stem Morphemes, based on A.M. Halpern, Yuma, IJAL 12 (1946) and IJAL 13 (1947).


Glover, Bonnie, 1977, Tolkapaya Demonstratives, m.s.


Hynes, Dell H., 1960, Lexicostatistics So Far, Current Anthropology 1:3-44.


Kaufman, Terence, and Alan Shusterman, English-Paipai Vocabulary, preliminary version, m.s.


———, 1972, Metathesis in Yuman Languages, ms.


Mixco, Mauricio, 1976, Glossary of Reconstructions, unpub. ms.

Munro, Pamela, 1976, From Existential to Copula: The History of Yuman BE, presented at the Symposium on Mechanisms of Syntactic Change, UCSB.


____, 1966, Walapai II: Morphology, IJAL 32:2:141-63.


Sawyer, Jesse O, 1962, 100-Word Diagnostic List in Weppo, dittoed.


_____, Yavapai Phonology, unpub. ms. n.d.


Number 5

Occasional Papers On Linguistics


SIU

Department of Linguistics
Southern Illinois University at Carbondale
OCCASIONAL PAPERS ON LINGUISTICS

Number 5

PROCEEDINGS

OF THE

1978 HOKAN LANGUAGES WORKSHOP

James E. Redden, Editor

Held at

University of California, San Diego

June 27-29, 1978

Department of Linguistics
Southern Illinois University
Carbondale, Illinois

Library of Congress Catalog

Number 79-38629
PREFACE

Unfortunately, everyone who presented a paper at the 1978 Hokan Languages Workshop was not able to prepare a final version for inclusion in this volume. All the papers in this volume were presented in an earlier version at the 1978 workshop. The papers are arranged in the order that they appeared on the program at the workshop.

The participants of the 1978 Hokan Languages Workshop gratefully acknowledge all the work done by Professor Carol Baker Slater and the students at the University of California, San Diego, which made the workshop run so smoothly and enjoyably.

Copies of the 1977 workshop are still available from the Department of Linguistics, Southern Illinois University, Carbondale, IL 62901. The volumes for the 1975 and 1976 workshops, which appeared in the SIU-C University Museum Studies, are now out of print, but copies may be obtained in microfiche or hard-bound copies from ERIC Clearinghouse on Languages and Linguistics, Center for Applied Linguistics, 1611 N. Kent Street, Arlington, VA 22209.

The 1979 Hokan Languages Workshop will be held at the University of California, Los Angeles, June 26-28. The proceedings of the 1979 workshop will appear in Occasional Papers On Linguistics in late spring 1980. Copies may be ordered from the Department of Linguistics, Southern Illinois University, Carbondale, IL 62901.

James E. Redden
Carbondale, May 1979
CONTENTS

Webb, Nancy M.
The Semantic Domain 'tree': Hokan Lexical Evidence 1

Redden, James E.
Notes on Walapai Syntax II 8

Watahomigic, Lucille J., Malinda Fowskey, and Akira Y. Yamamoto
The Structure of Nominal Modifiers 11

Hardy, Heather K.
An Integrated Account of the Morpheme $e$ in Tolkapaya 19

Hardy, Heather K.
The Development of the Pai Vowel System 29

Munro, Pamela
Reduplication in Mojave—and Yuman 42

Oswalt, Robert L.
An Exploration of the Affinity of Wappo and Some Hokan and Penutian Languages 56

Bendixen, Birgitte
Aspects of the Rhythmical Structure of Cocopa 72

Bibliography 91