Language and Region Codes for the Standard Cross-Cultural Sample

Michael L. Burton
University of California, Irvine

New language and region codes are developed for the Standard Cross-Cultural Sample. The region codes were developed previously with a different sample and tested against social structure data. The language codes incorporate information from recent publications on language history and are presented at multiple levels, providing information about the taxonomic relationships among languages.

Variation by region and language family has been an important topic in cross-cultural research. Regions have been used for replication of cross-cultural findings (Burton & White, 1991) and as the basis for stratified sampling (Murdock, 1967; Murdock & White, 1969), and language codes have been used in studies of Galton's

Author's Note: In working on this project, I benefited from discussions with and commentary by Cecil Brown, Stanley Witkowski, and Terrence Kaufman. Stanley Witkowski's comments on this project during his participation in the HRAF summer institute in comparative research methods led to the current multilevel coding.
problem (Burton & White, 1984; Dow, Burton, Reitz, & White, 1984; White, Burton, & Dow, 1981).

The most widely used cross-cultural codes for regions and language groups are those developed by Murdock (1967) for the Ethnographic Atlas. These are now more than 30 years old. Burton, Moore, Whiting, & Romney (1996) recently have proposed an improved set of regional categories, and there are several new publications in historical linguistics that can be used to formulate language codes (Campbell, 1997; Greenberg, 1987; Kaufman, 1994; Mosely & Asher, 1994; Ruhlen, 1991). Here, I have coded the 186 societies of the Standard Cross-Cultural Sample (Murdock & White, 1969) using the new regions and the various new studies of language families.

The revised regional classification was developed based on concepts of precapitalist world systems. There are two world systems: the Middle Old World in North Africa and Southern Eurasia, and the system of agrarian states in the New World that extended south from Mesoamerica to the Andes. These and the other regions are described briefly below and in more detail in Burton et al. (1996). The classification was tested using Murdock’s (1967, 1970) social structure data and found to fit the data much better than Murdock’s classification.

Language classifications seem to evoke strong feelings among comparative linguists. In fact, progress on this article was delayed by the process of asking comparative linguists to comment on the language classifications. In short, they do not agree. The problem often is described as the difference between lumpers and splitters, but the difference is better described in terms of methodology because both groups have the same goal—to classify languages into larger groupings. One group (the so-called “lumpers”) favors Greenberg’s (1963) method of mass comparisons, a statistical approach which, however, does not use particularly powerful methods of statistical analysis. The second group (the “splitters”) requires strong evidence for reconstruction of relationships between languages and rejects higher order groupings based on mass comparisons.

The greatest differences between the two approaches occur with classification of American Indian languages. Greenberg’s (1987) classification of these languages into three major groups is not widely accepted by specialists on American Indian languages,
whereas his earlier (Greenberg, 1963) classification of African languages still is closely followed.

Although I am sympathetic to statistical approaches and a great admirer of Greenberg's work, the lumper approach is less well suited to the purpose of this article, which is to provide information about language groupings upon which most historical linguists would agree. Information about some of the proposals for more macroscopic groupings appears at the end of the article.

A note is required here about the use of language codes to test Galton's problem. A shared language history is a plausible index of shared history, and in many (but not all) cases, language families provide useful information of a taxonomic nature. In working on Galton's problem, we used taxonomic relationships among languages as the basis for computing proximity measures among the various languages (Dow et al., 1984; White et al., 1981). We used these measures in analyses derived from spatial autocorrelation methods that we called language autocorrelation analysis. In studies of the gender division of labor, we found an autocorrelation effect that could be located within the group of societies with Bantu languages—an effect that we could detect but that was too small to affect the validity of the larger research project. As we shall see below, the Bantu languages are very closely related within a larger language family, so the main generalization from this work is that language autocorrelation is likely to occur only among very closely related languages. That would imply that information about very high-level groupings, such as Greenberg's Amerind language phyllum (Greenberg, 1987), would have little relevance to testing nomothetic cross-cultural hypotheses even if they were to be more widely accepted by historical linguists.

Previous language codes for cross-cultural research have focused on a single level of multilevel taxonomies. Here, I present several levels of language codes. Doing so provides more information about known relationships among languages and would allow other researchers to replicate the kind of autocorrelation analyses described above, which require the full taxonomic information.

Table 1 provides a brief description of each region and a tabulation of the number of standard sample societies that fall within the region. Table 2 presents the language codebook, with frequency counts for each language group. Following Campbell (1997), I do not use the term phyllum. The codes are organized by families (the
highest level groupings) and two levels of subfamilies. The actual codes are in Table 3.

The three levels coded allow most language families to be divided into small groups of standard sample societies, among which further differentiation should not be necessary. However, there remain two large groupings at the second subfamily level. These fall within two language families—Niger-Congo and Austronesian. Here, I discuss each of these in more detail.

**NIGER-CONGO**

In some classifications (e.g., Ruhlen, 1991), Niger-Congo and Kordofanian are combined into a larger group called Niger-
| Family          | Subfamilies          | N |
|-----------------|----------------------|---|
| 1 Khoisan       |                      | 3 |
| 1 Southern Khoisan |                    | 2 |
| 2 Hadza         |                      | 1 |
| 2 Niger-Congo   |                      | 23|
| 3 Central       |                      | 18|
| 4 Mande         |                      | 2 |
| 5 West Atlantic |                      | 2 |
| 6 Kordofanian   |                      | 1 |
| 3 Nilo-Saharan  |                      | 7 |
| 6 Songhai       |                      | 1 |
| 7 East Sudanic  |                      | 3 |
| 8 Fur           |                      | 1 |
| 9 Saharan       |                      | 1 |
| 10 Komuz        |                      | 1 |
| 4 Afro-Asiatic  |                      | 13|
| 11 Berber       |                      | 2 |
| 12 Chadic       |                      | 2 |
| 13 Omotic       |                      | 1 |
| 14 Cushitic     |                      | 3 |
| 15 Semitic      |                      | 5 |
| 16 Central      |                      | 1 |
| 17 Eastern      |                      | 2 |
| 18 Central      |                      | 4 |
| 19 Southern     |                      | 1 |
| 5 Indo-European |                      | 12|
| 20 Armenian     |                      | 1 |
| 21 Indo-Iranian |                      | 5 |
| 22 Albanian     |                      | 1 |
| 23 Italian      |                      | 1 |
| 24 Celtic       |                      | 1 |
| 25 Balto-Slavic |                      | 1 |
| 26 African-IE Creole |              | 2 |
| 6 Dravidian     |                      | 2 |
| 28 Central Dravidian |                | 1 |
| 29 South Dravidian |                  | 1 |

(continued)
TABLE 2 Continued

| Family       | Subfamilies       | N  |
|--------------|-------------------|----|
| 7            | Caucasian         | 1  |
| 8            | Sino-Tibetan      | 6  |
| 30           | Sinitic           | 1  |
| 31           | Tibeto-Burman     | 5  |
| 31M          | Tibetic           | 1  |
| 31M4         | Baric             | 1  |
| 31M5         | Burmic            | 3  |
| 9            | Austroasiatic     | 6  |
| 32           | Munda             | 1  |
| 33           | Mon-Khmer         | 5  |
| 33M6         | North             | 2  |
| 33M7         | East              | 1  |
| 33M8         | South             | 2  |
| 10           | Daic              | 1  |
| 11           | Austronesian      | 25 |
| 34           | Atalyic           | 1  |
| 35           | Malay-Polynesian  | 24 |
| 35P1         | Western           | 11 |
| 35P2         | Central and Eastern | 13 |
| 12           | Andaman           | 1  |
| 13           | Trans New Guinea  | 4  |
| 36           | Main              | 2  |
| 37           | Trans Fly         | 1  |
| 38           | Timor-Alor        | 1  |
| 14           | West Papuan       | 1  |
| 15           | Sepik-Ramu        | 1  |
| 16           | East Papuan       | 1  |
| 17           | Australian        | 2  |
| 20           | Uralic-Yukaghir   | 3  |
| 40           | Yukaghir          | 1  |
| 41           | Uralic            | 2  |
| 21           | Altaic            | 4  |
| 42           | Turkic            | 2  |
| 43           | Mongolian-Tungus  | 2  |
| 22           | Chukchi-Kamchatkan| 1  |
| 30           | Eskimo-Aleut      | 2  |
| 44           | Aleut             | 1  |
| 45           | Eskimo            | 1  |
| 31           | Na-Dene           | 6  |
| 50           | Haida             | 1  |
| 51           | Athabaskan-Eyak   | 5  |
| 51N1         | Eyak              | 1  |
| 51N2         | Athabaskan        | 4  |
### TABLE 2 Continued

| Family | Subfamilies | N  |
|--------|-------------|----|
| 32     | Algic       | 5  |
| 52     | Ritwan      | 1  |
| 53     | Algonquian  | 4  |
| 33     | Salish      | 2  |
| 34     | Siouan      | 2  |
| 35     | Iroquian    | 1  |
| 36     | Caddoan     | 1  |
| 37     | Hokan       | 2  |
| 38     | California and plateau Penutian | 2 |
| 39     | Natchez-Muskogean | 2 |
| 40     | Uto-Aztecan  | 5  |
| 54     | Northern    | 2  |
| 55     | Southern    | 3  |
|        | N3 Pimic    | 1  |
|        | N4 Corachol-Aztecan | 2 |
| 41     | Macro-Mayan | 2  |
| 56     | Mixe-Zoquean| 1  |
| 57     | Mayan       | 1  |
| 42     | Chibcha-Misumalpan | 3 |
| 58     | Misumalpan  | 1  |
| 59     | Chibchan    | 2  |
| 43     | Macro-Arawakan | 3 |
| 60     | Maiipuruan (Arawakan) | 2 |
| 61     | Otomakoan   | 1  |
| 44     | Macro-Paezan | 2 |
| 45     | Macro-Panoan | 2 |
| 62     | Panoan      | 1  |
| 46     | Tupi-Carib  | 5  |
| 63     | Carib       | 1  |
| 64     | Tupian      | 4  |
|        | S1 Tupi-Guarani | 3 |
|        | S2 Munduruku | 1 |
| 47     | Tucanoan    | 1  |
| 48     | Jivaroan    | 1  |
| 49     | Quechumaran  | 2 |
| 50     | Nambiquaran  | 1  |
| 51     | Macro Ge    | 4  |
| 65     | Botocudan   | 1  |
| 66     | Ge          | 3  |
| 52     | Mascoian    | 1  |
| 53     | Guaykuruayan| 1  |
| 54     | Araucanian  | 1  |
| 55     | Tehuelche   | 1  |
| 99     | Isolate     | 9  |
### TABLE 3
Language and Region Codes

| Identification Number | Name               | Region | Family | Subfamily 1 | Subfamily 2 |
|-----------------------|--------------------|--------|--------|-------------|-------------|
| 001                   | Nama               | 1      | 01     | 01          |             |
| 002                   | !Kung              | 1      | 01     | 01          |             |
| 003                   | Thonga             | 1      | 02     | 03          | A2          |
| 004                   | Lozi               | 1      | 02     | 03          | A2          |
| 005                   | Mbundu             | 1      | 02     | 03          | A2          |
| 006                   | Suku               | 1      | 02     | 03          | A2          |
| 007                   | Bemba              | 1      | 02     | 03          | A2          |
| 008                   | Nyakyusa           | 1      | 02     | 03          | A2          |
| 009                   | Hadza              | 1      | 01     | 02          |             |
| 010                   | Luguru             | 1      | 02     | 03          | A2          |
| 011                   | Gikuyu             | 1      | 02     | 03          | A2          |
| 012                   | Ganda              | 1      | 02     | 03          | A2          |
| 013                   | Mbuti              | 1      | 02     | 03          | A2          |
| 014                   | Nkundo Mongo       | 1      | 02     | 03          | A2          |
| 015                   | Banen              | 1      | 02     | 03          | A2          |
| 016                   | Tiv                | 1      | 02     | 03          | A2          |
| 017                   | Ibo                | 1      | 02     | 03          | A2          |
| 018                   | Fon                | 1      | 02     | 03          | A2          |
| 019                   | Ashanti            | 1      | 02     | 03          | A2          |
| 020                   | Mende              | 1      | 02     | 04          |             |
| 021                   | Wolof              | 1      | 02     | 05          |             |
| 022                   | Bambara            | 1      | 02     | 04          |             |
| 023                   | Tallensi           | 1      | 02     | 03          | A1          |
| 024                   | Songhai            | 1      | 03     | 06          |             |
| 025                   | Wodaabe Fulani     | 1      | 02     | 05          |             |
| 026                   | Hausa              | 1      | 04     | 12          |             |
| 027                   | Massa              | 1      | 04     | 12          |             |
| 028                   | Azande             | 1      | 02     | 03          | A1          |
| 029                   | Fur                | 1      | 03     | 08          |             |
| 030                   | Otoro Nuba         | 1      | 02     | 06          |             |
| 031                   | Shilluk            | 1      | 03     | 07          | A4          |
| 032                   | Mao                | 1      | 03     | 10          |             |
| 033                   | Kaffa              | 1      | 04     | 13          |             |
| 034                   | Maasai             | 1      | 03     | 07          | A4          |
| 035                   | Konso              | 1      | 04     | 14          | A7          |
| 036                   | Somali             | 1      | 04     | 14          | A7          |
| 037                   | Amhara             | 2      | 04     | 15          | A9          |
| 038                   | Bogo               | 2      | 04     | 14          | A6          |
| 039                   | Kenuzi Nubian      | 2      | 03     | 07          | A5          |
| 040                   | Teda               | 2      | 03     | 09          |             |
| 041                   | Tuareg             | 2      | 04     | 11          |             |
| 042                   | Riffian            | 2      | 04     | 11          |             |
### TABLE 3 Continued

| Identification Number | Name            | Region | Family | Subfamily 1 | Subfamily 2 |
|-----------------------|-----------------|--------|--------|-------------|-------------|
| 043                   | Egyptian        | 2      | 04     | 15          | A8          |
| 044                   | Hebrew          | 2      | 04     | 15          | A8          |
| 045                   | Babylonian      | 2      | 04     | 15          | A8          |
| 046                   | R calam Bedouin | 2      | 04     | 15          | A8          |
| 047                   | Turks           | 2      | 21     | 42          |             |
| 048                   | Gheg Albanian   | 5      | 05     | 22          |             |
| 049                   | Roman           | 5      | 05     | 23          |             |
| 050                   | Basque          | 5      | 99     |             |             |
| 051                   | Irish           | 5      | 05     | 24          |             |
| 052                   | Lapp (Saami)    | 5      | 20     | 41          |             |
| 053                   | Nenets (Samoyed)| 5      | 20     | 41          |             |
| 054                   | Russian         | 5      | 05     | 25          |             |
| 055                   | Abkhaz          | 2      | 07     |             |             |
| 056                   | Armenian        | 2      | 05     | 20          |             |
| 057                   | Kurd            | 2      | 05     | 21          | M2          |
| 058                   | Basseri         | 2      | 05     | 21          | M2          |
| 059                   | West Punjabi    | 2      | 05     | 21          | M1          |
| 060                   | Gond            | 2      | 06     | 28          |             |
| 061                   | Toda            | 2      | 06     | 29          |             |
| 062                   | Santal          | 2      | 09     | 32          |             |
| 063                   | Uttarakhand     | 2      | 05     | 21          | M1          |
| 064                   | Burusho         | 2      | 99     |             |             |
| 065                   | Kazak           | 2      | 21     | 42          |             |
| 066                   | Khalka Mongols  | 2      | 21     | 43          |             |
| 067                   | Lolo            | 2      | 08     | 31          | M5          |
| 068                   | Lepcha          | 2      | 08     | 31          | M3          |
| 069                   | Garo            | 2      | 08     | 31          | M4          |
| 070                   | Lakhin          | 2      | 08     | 31          | M5          |
| 071                   | Burmese         | 3      | 08     | 31          | M5          |
| 072                   | Lamet           | 3      | 09     | 33          | M6          |
| 073                   | Vietnamease      | 2      | 09     | 33          | M6          |
| 074                   | Rhade           | 2      | 11     | 35          | P1          |
| 075                   | Khmer           | 3      | 09     | 33          | M7          |
| 076                   | Central Thai     | 3      | 10     |             |             |
| 077                   | Semang          | 3      | 09     | 33          | M8          |
| 078                   | Nicobarese       | 3      | 09     | 33          | M8          |
| 079                   | Andamese        | 3      | 12     |             |             |
| 080                   | Vedda           | 2      | 05     | 21          | M1          |
| 081                   | Tanala          | 1      | 11     | 35          | P1          |
| 082                   | Negri Sembilan  | 3      | 11     | 35          | P1          |
| 083                   | Java            | 3      | 11     | 35          | P1          |
| 084                   | Bali            | 3      | 11     | 35          | P1          |

(continued)
| Identification Number | Name       | Region | Family | Subfamily 1 | Subfamily 2 |
|-----------------------|------------|--------|--------|-------------|-------------|
| 085                   | Iban       | 3      | 11     | 35          | P1          |
| 086                   | Badjau     | 3      | 11     | 35          | P1          |
| 087                   | Toradja    | 3      | 11     | 35          | P1          |
| 088                   | Tobelorese | 3      | 14     |             |             |
| 089                   | Alorese    | 3      | 13     | 38          |             |
| 090                   | Tiwi       | 4      | 17     |             |             |
| 091                   | Aranda     | 4      | 17     |             |             |
| 092                   | Orokaiva   | 4      | 13     | 36          |             |
| 093                   | Kimam      | 4      | 13     | 37          |             |
| 094                   | Kapauku    | 4      | 13     | 36          |             |
| 095                   | Kwoma      | 4      | 15     |             |             |
| 096                   | Manus      | 4      | 11     | 35          | P2          |
| 097                   | New Ireland| 4      | 11     | 35          | P2          |
| 098                   | Trobriand  | 4      | 11     | 35          | P2          |
| 099                   | Siual      | 4      | 16     |             |             |
| 100                   | Tikopia    | 4      | 11     | 35          | P2          |
| 101                   | Pentecost  | 4      | 11     | 35          | P2          |
| 102                   | Mbau Fijian| 4      | 11     | 35          | P2          |
| 103                   | Ajie       | 4      | 11     | 35          | P2          |
| 104                   | Maori      | 3      | 11     | 35          | P2          |
| 105                   | Marquesan  | 3      | 11     | 35          | P2          |
| 106                   | Samoan     | 3      | 11     | 35          | P2          |
| 107                   | Kiribati   | 3      | 11     | 35          | P2          |
| 108                   | Marshallese| 3      | 11     | 35          | P2          |
| 109                   | Chuuk (Truk)| 3    | 11     | 35          | P2          |
| 110                   | Yap        | 3      | 11     | 35          | P1          |
| 111                   | Palau      | 3      | 11     | 35          | P1          |
| 112                   | Ifugao     | 3      | 11     | 35          | P1          |
| 113                   | Atayal     | 3      | 11     | 34          |             |
| 114                   | Chinese    | 2      | 08     | 30          |             |
| 115                   | Manchu     | 2      | 21     | 43          |             |
| 116                   | Korean     | 5      | 99     |             |             |
| 117                   | Japanese   | 5      | 99     |             |             |
| 118                   | Ainu       | 5      | 99     |             |             |
| 119                   | Gilyak     | 5      | 99     |             |             |
| 120                   | Yukaghir   | 5      | 20     | 40          |             |
| 121                   | Chukchee   | 5      | 22     |             |             |
| 122                   | Ingallk   | 6      | 31     | 51          | N2          |
| 123                   | Aleut      | 5      | 30     | 44          |             |
| 124                   | Copper Eskimo | 5 | 30 | 45 | |
| 125                   | Montagnais | 7      | 32     | 53          |             |
| 126                   | Micmac     | 7      | 32     | 53          |             |
| Identification Number | Name      | Region | Family | Subfamily 1 | Subfamily 2 |
|-----------------------|-----------|--------|--------|-------------|-------------|
| 127                   | Salteaux  | 7      | 32     | 53          |             |
| 128                   | Slave     | 7      | 31     | 51          | N2          |
| 129                   | Kaska     | 6      | 31     | 51          | N2          |
| 130                   | Eyak      | 6      | 31     | 51          |             |
| 131                   | Haida     | 6      | 31     | 50          |             |
| 132                   | Bella Coola | 6      | 33     |             |             |
| 133                   | Twana     | 6      | 33     |             |             |
| 134                   | Yurok     | 7      | 32     | 52          |             |
| 135                   | Pomo      | 7      | 37     |             |             |
| 136                   | Yokuts    | 7      | 38     |             |             |
| 137                   | Paiute    | 7      | 40     | 54          |             |
| 138                   | Klamath   | 7      | 38     |             |             |
| 139                   | Kutenai   | 7      | 99     |             |             |
| 140                   | Gros Ventre | 7     | 32     | 53          |             |
| 141                   | Hidatsa   | 8      | 34     |             |             |
| 142                   | Pawnee    | 8      | 36     |             |             |
| 143                   | Omaha     | 8      | 34     |             |             |
| 144                   | Huron     | 8      | 35     |             |             |
| 145                   | Creek     | 8      | 39     |             |             |
| 146                   | Natchez   | 8      | 39     |             |             |
| 147                   | Comanche | 8      | 40     | 54          |             |
| 148                   | Chiricahua | 7      | 31     | 51          | N2          |
| 149                   | Zuni      | 9      | 99     |             |             |
| 150                   | Havasupai | 7      | 37     |             |             |
| 151                   | Papago    | 9      | 40     | 55          | N3          |
| 152                   | Huichol   | 9      | 40     | 55          | N4          |
| 153                   | Aztec     | 9      | 40     | 55          | N4          |
| 154                   | Sierra Popoluca | 9    | 41     | 56          |             |
| 155                   | Quiche    | 9      | 41     | 57          |             |
| 156                   | Miskito   | 9      | 42     | 58          |             |
| 157                   | Bribri    | 9      | 42     | 59          |             |
| 158                   | Cuna      | 9      | 42     | 59          |             |
| 159                   | Goajiro   | 9      | 43     | 60          |             |
| 160                   | Haitian   | 8      | 05     | 26          |             |
| 161                   | Callinago | 8      | 43     | 60          |             |
| 162                   | Warrau    | 8      | 44     |             |             |
| 163                   | Yanomamo  | 8      | 45     |             |             |
| 164                   | Barama Carib | 8    | 46     | 63          |             |
| 165                   | Saramacca | 8      | 05     | 26          |             |
| 166                   | Mundurucu | 8      | 46     | 64          | S2          |
| 167                   | Cubeo     | 8      | 47     |             |             |
| 168                   | Cayapa    | 9      | 44     |             |             |

(continued)
Kordofanian. However, following Wald (1994), I have included Kordofanian as one of the subgroups of Niger-Congo, at the same level as Mande. This produces three main subfamilies that include societies from the Standard Cross-Cultural Sample—Mande, West Atlantic, and Central Niger-Congo. The latter group includes 18 societies from the Standard Cross-Cultural Sample. Figure 1 presents a simplified taxonomy of the Central Niger-Congo family, listing only those groups that are necessary to distinguish among these 18 languages. Of the 18 languages, 12 are Bantu languages and 13 are in the Bantoid group. It takes seven levels of the taxonomy of Central Niger-Congo languages to get to Bantu and two more to get to Central Bantu (see Figure 1), with 10 members (more than 5% of the standard sample). The depth of this group within the taxonomy shows its relatively shallow time depth. The Bantu languages are the largest group in the sample with such a close historical relationship, so it is not surprising that they were the basis for our finding of language autocorrelation, described above.
1. North Central (2)
   Azande
   Tallensi

2. South Central (16)
   a. Western (2)
      Ashanti
      Fon
   b. Eastern (14)
      i. Lower Niger (1)
         Igbo
      ii. Benue Zambesi (13)
         a. Cara
         b. Nyima
            i. Plateau
            ii. Wel
               1. Bendi-Bokyi
               2. Bantoid
                  a. Non-Bantu (1)
                     Tiv
                  b. Broad Bantu (12)
                     i. Bane (1)
                        Banen
                     ii. Narrow Bantu (11)
                        a. Northwest Bantu (1)
                           Nkundo Mondo
                        b. Central Bantu (10)
                           Thonga
                           Lozi
                           Mbundu
                           Suku
                           Bemba
                           Nyakyusa
                           Luguru
                           Gikuyu
                           Ganda
                           Mbuti

Figure 1: Central Niger-Congo Languages (N = 18)

AUSTRONESIAN

Austronesian (see Figure 2) includes 25 members in the Standard Cross-Cultural Sample. Four levels down, 13 of the languages are in the Oceanic branch of Eastern Malayo-Polynesian,
1. Ataylic (1)
   Atayal
2 Malayo-Polynesian (24)
   a Western Malayo-Polynesian (11)
      i Sundic (5)
         a Malayic (3)
            Rhade
            Negri-Sembilan
            Iban
         b J ava (1)
            J ava
         c Bali-Susak (1)
            Bali
      ii Borneo (1)
         Tanala
      iii Samu-Bajaw (1)
         Badjau
   iv Celebes (1)
      Toradja
   v North Philippines (1)
      Ifugao
   vi Yap (1)
      Yap
   vii Palau (1)
      Palau
   b Central and Eastern Malayo-Polynesian (13)
      i Central Malayo-Polynesian (0)
      ii Eastern Malayo-Polynesian (13)
         a South Halmahera-Northwest New Guinea (0)
         b Oceanic (13)
            1 Admiralty (1)
               Manus
            2 New Ireland-Tolai (1)
               New Ireland
            3 Milne Bay (1)
               Trobriands
            4 New Hebrides (1)
               Aije
            5 Remote Oceanic (9)
               a Micronesia (3)
                  Kiribati
                  Marshallese
                  Chuuk
               b Central & New Hebrides (1)
                  Pentecost

Figure 2: Austronesian Language Family (N = 25)
which contains many Pacific Island languages. Here, the situation is similar to that with the Bantu languages, where a relatively cohesive group of people settled a large geographic area. It takes three more levels of the taxonomy to get to the four Polynesian languages.

CREOLE LANGUAGES

Taxonomic trees cannot capture the entire complexity of relationships among languages. For example, a taxonomic tree cannot accurately depict the known relationships of English, a Germanic language, with the Celtic and Romance languages; the relationships of Swahili both to Arabic and to the East African Bantu languages; or the language mixing that has occurred between some Austronesian and non-Austronesian languages in Melanesia. If language is to be used as a proxy for long-term historical relationships, one must be mindful of other kinds of processes that may affect the accuracy of taxonomic representations. For example, the Mbuti speak a Bantu language, but this does not index long-term historical linkages with other Bantu-speaking societies.

Two societies in the Standard Cross-Cultural Sample have Creole languages—the Haitians and the Saramacca. Both have strong influences from Indo-European and African languages. I have added an extra category under Indo-European to include these two.
In the Old World (Africa, Eurasia, and the Pacific), there is a strong relationship between regions and language families (see Table 4). This is represented visually in Figure 3 with a correspondence analysis. Here, we can see three lines of language families: (a) African language families, (b) language families of North Eurasia and Circumpolar, and (c) language families of Southeast Asia and the Pacific. The three lines are connected through the Middle Old World, the center of the Old World system.

| Language Family          | Middle Africa | South East Asia and New Guinea, Insular Pacific | North and Australia, and Melanesia | Circum-Polar |
|--------------------------|---------------|-----------------------------------------------|-----------------------------------|--------------|
| Khoisan                  | 3             |                                               |                                   |              |
| Niger-Congo              | 23            |                                               |                                   |              |
| Nilo-Saharan             | 5             | 2                                             |                                   |              |
| Afro-Asiatic             | 5             | 8                                             |                                   |              |
| Indo-European            | 6             |                                               |                                   | 4            |
| Dravidian                | 2             |                                               |                                   |              |
| Caucasian                | 1             |                                               |                                   |              |
| Altaic                   | 4             |                                               |                                   |              |
| Sino-Tibetan             | 5             | 1                                             |                                   |              |
| Austroasiatic            | 2             | 4                                             |                                   |              |
| Daic                     | 1             |                                               |                                   |              |
| Austronesian             | 1             | 1                                             | 16                                | 7            |
| Andaman                  | 1             |                                               |                                   |              |
| West Papuan              | 1             |                                               |                                   |              |
| Trans New Guinea         | 1             |                                               |                                   | 3            |
| Sepik-Ramu               | 1             |                                               |                                   |              |
| East Papuan              | 1             |                                               |                                   |              |
| Australian               | 2             |                                               |                                   |              |
| Uralic-Yukaghir          | 3             |                                               |                                   |              |
| Chukchi-Kamchatkan       |               |                                               |                                   | 1            |
| Eskimo-Aleut             |               |                                               |                                   | 2            |
Language families of the Americas show a different pattern of correspondence between regions and language families. Table 5 tabulates the many American language families against four American regions. The correspondence analysis of this table appears in Figure 4. Here, there are two main groupings, the first focussed in the west and northwest, the second including the Eastern Americas and Mesoamerica to the Andes.
PROPOSALS FOR MACROSCOPIC GROUPINGS

In addition to proposals discussed above, some scholars lump language families 13 to 16 into a single Papuan phylum. Korean and Japanese sometimes are included in a single family, which is often included in the Altaic language family. Although those proposals are open to debate, the earlier proposal for linking Uralic and Altaic has fallen into disuse.

I have included Haida in NaDene, following Campbell (1997), but this is not universally accepted. My usages of Penutian and Hokan follow Campbell’s more conservative approach, and these are much smaller groupings than the larger groupings that originally were proposed by Sapir (1921).

| Language Family | Northwest Coast | North and West | East | Mesoamerica, Andes |
|-----------------|-----------------|----------------|------|---------------------|
| Salish          | 2               |                |      |                     |
| NaDene          | 4               | 2              |      |                     |
| Algic           |                 | 5              |      |                     |
| Hokan           |                 | 2              |      |                     |
| Penutian        |                 | 2              |      |                     |
| Siouan          |                 | 2              |      |                     |
| Iroquian        |                 | 1              |      |                     |
| Caddoan         |                 | 1              |      |                     |
| Natchez         |                 | 2              |      |                     |
| Panoan          |                 | 2              |      |                     |
| Tupian          |                 | 4              |      |                     |
| Tucanoan        |                 | 2              |      |                     |
| Jivarano        |                 | 1              |      |                     |
| Nambiquara      |                 | 1              |      |                     |
| Macro Ge        |                 | 4              |      |                     |
| Mascoian        |                 | 1              |      |                     |
| IE Creole       |                 | 2              |      |                     |
| Guaykuran       |                 | 1              |      |                     |
| Uto-Aztecan     | 1               | 1              | 3    |                     |
| Paezan          | 1               | 1              |      |                     |
| Arawakan        | 2               | 1              |      |                     |
| Mayan           | 2               |                |      |                     |
| Chibchan        | 3               |                |      |                     |
| Quechumara      | 2               |                |      |                     |
Notes

1. The language codes presented here are not the same as the codes published earlier in the electronic journal World Cultures. Those codes were based on Voegelin and Voegelin (1977).

2. This and the Austronesian taxonomy below are simplified by deletion of branches that do not contain standard sample societies. The full taxonomies are much more complicated.

3. Benedict (1942, 1975) proposed a link between Austronesian and the Thai (Daic) languages in a group called Austro-Thai. This is accepted by many but not all historical linguistics. Less support is given to an even
more macroscopic grouping of Austro-Thai with Austroasiatic in a phylum called Austric, described in Ruhlen (1991).
4. Three Far South societies not included.

References

Benedict, P. (1942). Thai, Kadai, and Indonesian: A new alignment in Southeastern Asia. American Anthropologist, 44, 576-601.
Benedict, P. (1975). Austro-Thai: Language and culture. New Haven, CT: HRAF.
Burton, M. L., Moore, C. C., Whiting, J. W. M., & Romney, A. K. (1996). Regions based on social structure. Current Anthropology, 37, 87-123.
Burton, M. L., & White, D. R. (1984). Sexual division of labor in agriculture. American Anthropologist, 86, 568-583.
Burton, M. L., & White, D. R. (1991). Regional comparisons, replications, and historical network analysis. Behavior Science Research, 25, 55-78.
Campbell, L. (1997). American Indian Languages: The historical linguistics of Native America. New York: Oxford University Press.
Dow, M. M., Burton, M. L., Reitz, K., & White, D. R. (1984). Galton’s problem as network autocorrelation. American Ethnologist, 11, 754-770.
Driver, H. E. (1961). Indians of North America. Chicago: University of Chicago Press.
Greenberg, J. H. (1963). The languages of Africa. Bloomington: University of Indiana Press.
Greenberg, J. H. (1987). Language in the Americas. Stanford, CA: Stanford University Press.
Kaufman, T. (1994). The native languages of South America. In C. Moseley & R. E. Asher (Eds.), Atlas of the world’s languages (pp. 46-70). Boston: Routledge Kegan Paul.
Mosely, C., & Asher, R. E. (1994). Atlas of the world’s languages. Boston: Routledge Kegan Paul.
Murdock, G. P. (1967). Ethnographic atlas. Pittsburgh: University of Pittsburgh Press.
Murdock, G. P. (1970). Kin term pattern variables and their distribution. Ethnology, 8, 329-369.
Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. Ethnology, 8, 329-369.
Ruhlen, M. (1991). A guide to the world’s languages: Vol. 1. Classification. Stanford, CA: Stanford University Press.
Sapir, E. (1921). A bird’s-eye view of American languages north of Mexico. Science, 54, 408.
Voegelin, C. F., & Voegelin, F. M. (1977). Classification and index of the world’s languages. New York: Elsevier North-Holland.
Wald, B. (1994). Sub-Saharan Africa. In C. Moseley & R. E. Asher (Eds.), Atlas of the world’s languages (pp. 289-309). Boston: Routledge Kegan Paul.
White, D. R., Burton, M. L., & Dow, M. M. (1981). Sexual division of labor in African agriculture. American Anthropologist, 83, 824-849.

Michael L. Burton is professor of anthropology at the University of California, Irvine. He has done field research in Kenya and Micronesia and has published on cross-cultural research, cognitive anthropology, and gender division of labor. He currently is working on a comparative study of household economy in four Micronesian societies.