**Language classification system, based on Ruhlen**

John Simister

As part of processing survey data for my own analysis (for journal articles), I used the Ruhlen language classification system. I base it on Merritt Ruhlen (1987), ‘A guide to the world's languages, vol 1: Classification’ (Stanford CA, Stanford Univ Press, pp.301-78). Info file by Kazuo Morishima, Lawrence Reid, Saranya Savetamalya, David Stampe, Ryo Stanwood & Fumio Usami: 1988. HTML version by Stampe, November 2001 (stampe@hawaii.edu). I downloaded information from [www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html](http://www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html) on 15th January 2004, but that webpage is no longer available. I found (on 20th October 2014) that the ‘Wayback’ website has a backup copy: [http://web.archive.org/web/\*/www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html](http://web.archive.org/web/*/www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html)

e.g. for 31st March 2012 (later versions on the Wayback website don’t seem to work).

The Ruhlen system is not the only way to classify languages. Other classification systems include Ethnologue, as shown at <http://www.ethnologue.com/language/lee> I am not a linguistics expert, but I get the impression Ruhlen’s classification is similar to other systems such as Ethnologue. I converted the Ruhlen system to something I found easier to use, by creating numeric codes (e.g. 10302002.01); these numbers preserve the hierarchy which Ruhlen (and others) developed. I produced a spreadsheet ***[Carol: please put a hyperlink here]*** from information on the [www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html](http://www.ling.hawaii.edu/faculty/stampe/Linguistics/Ruhlen/ruhlen.html) webpages. There are many more languages in the spreadsheet than I include below: I only include languages I found in ‘Demographic and Health Survey’ (DHS) or ‘Work, Attitudes and Spending’ (WAS) surveys I have processed so far. I created a unique numeric code for each language based on my spreadsheet; dividing this by a number such as 100 or 10000 gives a language group, as shown in the SPSS syntax below (which I wrote).

I use language (or language group) as an indicator of ethnicity: in my opinion, the language a person speaks is usually a good guide to their ‘tribe’. I am a native English speaker, with white skin; I consider myself a member of the English tribe (sometimes called Anglo-Saxon). I think my values have much in common with a nearby ‘tribe’, Welsh-speakers; but there are some Welsh cultural practices such as their Eisteddfod which I know little about – see <http://www.eisteddfod.org.uk/english/>

John Simister, 20th October 2014.

\* Add more languages, if needed (search for the new language; then copy the cell from Excel to this file).

\* I shorten the labels in 'Ruhlen language classification values.xls' in this SPSS syntax file: I just put the specific

\* language, not the language group (e.g. Amerindian) because groups are labelled at the bottom of this file.

\* If there's a language used in DHS which seems to correspond to a similar language in Ruhlen, then I put the

\* DHS value label in brackets after the Ruhlen name.

\*.

val label INT\_LANG, INT2LANG, ethnic

20201000.01 "Bobo (Bobo-Fing)"

20201003.01 "Sarkolé/Soninké/Marka" 20201003.02 "Bozo/Somonon"

20201006.01 "Vai"

20201008.01 "Mandigo (Mandingo, Mandinka)"

20201009.01 "Malinké (Maninka)" 20201009.02 "Bambara" 20201009.03 "Dyula (Dioulali)"

20201010.02 "Lorma (Loma)"

20201010.03 "Kpelle"

20201011.01 "Mende"

20201011.02 "Bandi (Gbandi)"

20201013.01 "Mano" 20201013.02 "Gio (Dan)"

20202014.01 "Gola" 20201014.03 "Nwa"

20201015.02 "Busa (Shangawa; Wawa)"

20202016.04 "Kisi (Bullom-Kissi: Liberia)"

20202003.01 "Fulani (Fulfulde, Foulbé, Peulh, Bororo)"

20202003.03 "Wolof"

20202013.02 "Limba"

20202016.01 "Bom (Boum)"

20202019.02 "Kuwaa (Belle)"

20202023.02 "Kru (Krumen)"

20202023.03 "Grebo"

20202025.01 "Volta-Congo [Kru]: Bassa (Liberia)"

20202025.02 "Dey (Dewoin)"

20202026.01 "Krahn"

20202027.01 "Guere (Gueré)"

20202028.01 "Dogon (Nanga)"

20202032.01 "Senari (Senoufo)"

20202034.02 "Sénoufo/Minianka (Mianka)"

20202044.03 "More (Moore; Mossi)"

20202044.11 "Dagbani"

20202046.07 "Gurma (Gourmantche)"

20202051.00 "GURUNSI (Gouroussi)"

20202058.01 "Lobi"

20202060.00 "ADAMAWA (Kwa)"

20202060.02 "Longuda (Lunguda)"

20202060.03 "Fali (Fale, Kangou)"

20202060.04 "Nimbari (Samba, Ndi)"

20202061.01 "Awak"

20202061.04 "Tula"

20202061.05 "Waja"

20202063.01 "Chamba"

20202063.03 "Wom"

20202065.01 "Doyãyo (Doayo)"

20202065.03 "Duru (Dourou, Dii)"

20202065.06 "Kutin (Pere)"

20202065.09 "Vere (Verre)"

20202067.01 "Mumuye (Gongla)"

20202067.04 "Teme"

20202067.05 "Waka"

20202068.01 "Yendang (Poli, Bali)"

20202069.04 "Tupuri (Toupouri, Mambai)"

20202070.01 "Mbum (Mboum, Kepere)"

20202071.00 "Mundang-Tuburi-Mbum"

20202071.01 "Mundang (Moundang) (includes Kebbi)"

20202073.01 "Libo"

20202073.02 "Mboi"

20202074.01 "Yungur"

20202075.01 "Bambuka"

20202075.02 "Jen (Lo)"

20202075.04 "Munga"

20202076.00 "Bua"

20202077.00 "UBANGIAN (Ubangi and Itimbiri)"

20202079.01 "Gbaya (Baya)"

20202081.02 "Sango"

20202093.01 "Baka (Pygmee, Pygmy)"

20202102.02 "Golo"

20202110.00 "Ijo (Ijaw; Izon)"

20202110.01 "Kalabari"

20202110.02 "Nembe"

20202116.01 "Akan"

20202118.01 "Nzema"

20202124.01 "Gã"

20202126.01 "Ewe"

20202126.02 "Fon"

20202133.01 "Ebira (Ibira)"

20202134.01 "Gade"

20202135.01 "Nupe (Kakanda; Kupa)"

20202135.02 "Gwari (Gbaju, Gbagi)"

20202136.01 "Alago"

20202136.02 "Idoma"

20202136.03 "Yala (Yala Obubra)"

20202138.01 "Yoruba (& Itsekiri)"

20202138.02 "Igala/Igbala"

20202139.00 "Edo"

20202141.02 "Urhobo"

20202141.04 "Isoko"

20202142.02 "Esan"

20202142.04 "Iyekhee (Auchi)"

20202144.01 "Ekpeye"

20202145.02 "Igbo (=Ibo)"

20202145.06 "Ikwere"

20202147.02 "Kuteb (Akum)"

20202150.03 "Jukun"

20202151.01 "Bandawa"

20202154.00 "OGONI (Ogori)"

20202155.01 "Obolo (Oron)"

20202156.01 "Efik"

20202156.02 "Ibibio"

20202156.03 "Anaang (Annang)"

20202159.01 "Mbembe (Tigon)"

20202159.05 "Loko (Lukaa; Ugep)"

20202165.00 "CHAWAI (Butawa)"

20202165.01 "Piti"

20202166.01 "Kuzamani (Lere)"

20202166.03 "Rumaya"

20202166.06 "Kono"

20202166.08 "Kaivi (Kaibi)"

20202166.11 "Kinuku"

20202166.13 "Gure (Kahugu)"

20202168.01 "Reshe (Gungawa)"

20202167.02 "Butu (Kudawa)"

20202167.04 "Taura (Gana Gana)"

20202167.06 "Janji"

20202167.08 "Buji"

20202169.02 "Lopa (Lopawa)"

20202171.00 "KAMBARI (Tsikimba)"

20202173.04 "Pongu (Pangu)"

20202173.07 "Bassa-Kwomu (Bassa Komo)"

20202174.01 "Lela (Dakarkari, Badakare)"

20202178.01 "Birom"

20202179.01 "Katab (Kagoro; Kataf)"

20202180.03 "Koro"

20202181.02 "Kagoma"

20202181.03 "Jaba"

20202181.05 "Izarek (Jari)"

20202181.06 "Kaje"

20202182.02 "Kuturmi"

20202183.01 "Rukuba"

20202185.01 "Ninzam (Ninzo)"

20202185.02 "Mada"

20202186.01 "Eggon"

20202186.03 "Ake"

20202187.01 "Fyam (Pyem)"

20202188.01 "Tarok (Yergam)"

20202188.02 "Bashar (Basharawa)"

20202191.01 "Bokyi (Boki; Wula)"

20202192.02 "Bendi"

20202195.01 "Mambila"

20202195.08 "Vute (Baboute)"

20202196.01 "Tiv"

20202196.04 "Iceve-Maci (Icheve, Akwaya)"

20202200.00 "JARAWAN"

20202201.02 "Mbula"

20202201.07 "Kulung"

20202201.03 "Mama (Mama Plateau)"

20202203.00 "EKOID (Ekoi)"

20202203.02 "Ejagham (Ejaghan)"

20202204.03 "Kenyang (Banyangué/Bayangi)"

20202204.06 "Menka"

20202207.00 "NKAMBE"

20202207.08 "Yamba (Kaka)"

20202210.00 "NGEMBA (Nguemba, bamenda)"

20202210.04 "Nkwen (Nkwain/Bamendankwain)"

20202210.03 "Bafut"

20202210.07 "Awing"

20202211.00 "Bamileké"

20202211.13 "Bangwa (lebialem)"

20202211.15 "Bacham"

20202209.01 "Bamun (Bamoun)"

20202209.04 "Baba (Babar/Babare)"

20202210.01 "Mankon"

20202210.05 "Bambui (Mbatu)"

20202213.00 "RING"

20202215.02 "Isu"

20202216.01 "Babanki"

20202216.02 "Kom (Bikom)"

20202216.04 "Oku"

20202217.01 "Lamnso (Banso, Nso)"

20202218.00 "NDOP"

20202218.05 "Bamali"

20202220.00 "MOMO (Bameta/batibo/Meta, Ngamambo/Mbei)"

20202220.01 "Tadkon (Widikum)"

20202220.02 "Ngie"

20202222.00 "MENCHUM"

20202224.00 "Beboid: western (Mbu)"

20202225.01 "Bebe"

20202225.04 "Noni"

20202226.01 "Tikar (Tikari)"

20202230.00 "LUNDU-BALONG (Mbô, Bakaka, Akoose, Bakossi, Oroko/Bakundu/Batanga)"

20202230.03 "Balong"

20202231.02 "Kwiri (Bakweri)"

20202231.04 "Duala"

20202231.07 "Limba (Malimba)"

20202232.03 "Yasa (Yassa)"

20202233.02 "Bankon (Abo)"

20202233.03 "Basaa (Bassa); Basso'o"

20202233.04 "Tunen (Banen, Banen-Bandem)"

20202234.00 "BAFIA"

20202235.02 "Yambasa (Yambassa)"

20202236.00 "Ewondo (Beti)"

20202236.03 "Bebele (Bebil, Bobilis)"

20202236.04 "Bulu (Boulou)"

20202236.05 "Fang (Ntoumou)"

20202237.01 "Ngumba (Kwasio, Ngoumba)"

20202237.03 "Makaa (Maka, Meka, BIKelé, Nzime)"

20202237.05 "Konabem (Kounabembe, Mpongmpong)"

20202238.00 "KAKO"

20202238.02 "Pol"

20202266.02 "Kisii"

20202266.03 "Kuria"

20202267.01 "Kikuyu"

20202246.06 "Bali"

20202247.04 "Yanzi (Boma)"

20202252.06 "Lingala (Losengo)"

20202255.00 "MONGO ('Uele Lake Albert')"

20202267.02 "Embu"

20202267.03 "Meru"

20202267.05 "Kamba"

20202269.02 "Mijikenda(Nyika/Kauma/Ribe/Jibaba/Kambe/Duruma/Rabai)"

20202269.03 "Digo"

20202269.04 "Taita"

20202276.01 "Taveta"

20202278.02 "Swahili (Kiswahili including Bajuni; in DRC, Basele-K, Man. & Kivu)"

20202282.06 "Kongo, Koongo (Bakongo North and South)"

20202287.01 "Nyoro (Runyoro, Rutoro)" 20202287.02 "Luganda"

20202289.02 "Luhya"

20202292.01 "Kinyarwanda (Rwanda)"

20202295.04 "Lwena (Luvale)"

20202296.02 "Ndembo (Lunda)"

20202304.01 "Luba-Kasai, Tshiluba (Kasai, Katanga, Tanganika; Baluba)"

20202304.03 "Kiluba, Luba-Shaba, Luba-Katanga (Bas-Kasai & Kwilu-Kwngo)"

20202305.01 "Kaonde"

20202310.01 "Nyakyusa (Nkonde)"

20202311.02 "Bemba"

20202312.02 "Lala"

20202313.04 "Tonga (in Lenje-Tonga group)"

20202316.01 "Tumbuka"

20202317.01 "Chichewa (Chewa, Nyanja, Amanganja/Anyanja)"

20202318.04 "Sena"

20202321.01 "Yao"

20202322.02 "Lomwe"

20202328.01 "Shona"

20202328.04 "Karanga (Laka)"

20202329.01 "Venda"

20202330.01 "Tswana"

20202330.02 "Northern Sotho"

20202330.03 "Southern Sotho"

20202330.04 "Lozi"

20202331.01 "Xhosa"

20202331.02 "Zulu"

20202331.03 "Swazi"

20202331.04 "Ndebele"

20202331.05 "Ngoni"

20202332.03 "Tsonga/Shangaan"

30100000.01 "Sonrai (Songhai)"

30201001.00 "Kanuri (& Kanembu-Zaghawa; Kole, Sirata)"

30201002.01 "Tedaga (Tubu; Dazaga; Goran: Gorane)"

30300000.01 "Mimi"

30301000.00 "Maba (Mabang; Ouaddai)"

30400000.00 "Fur"

30502004.01 "Tama"

30502005.00 "Daju (1 of 15 Hadjarai groups)"

30503006.01 "Luo"

30503017.00 "Samburu"

30503017.02 "Maasai (Masai)"

30503018.01 "Ateso-Karamojong"

30503018.03 "Turkana"

30503018.04 "Teso"

30503020.00 "Kalenjin"

30503021.01 "Nandi"

30503021.02 "Kipsigis"

30503021.03 "Elgeyo"

30503021.04 "Tugen"

30503021.05 "Marakwet"

30503022.00 "Sabaot"

30503025.01 "Pokot"

30601007.01 "Sara-Bongo-Bagirmi (Gambai, Ngambaye/ Ngambai)"

30601007.05 "Sara"

30601008.02 "Bagirmi (Baguirmien)"

30602003.04 "Lugbara (Logbara)"

40200000.00 "BERBER"

40203005.02 "Tanachek (Tamasheq; Tuareg)"

40301000.01 "Masana (Masa, Massa, Walia)"

40301000.04 "Marba/Tanjile"

40302003.01 "Somrai"

40302008.01 "Kwanga/Kera"

40303003.01 "Tera"

40303004.01 "Ga'anda"

40303004.02 "Hona"

40303008.01 "Bura (Babur)"

40303009.01 "Margi"

40303009.02 "Kilba"

40303010.01 "Higi (Higgi, Kapsiki)"

40303010.02 "Bana (Guili)"

40303011.01 "Sukur"

40303012.01 "Lamang (Wagga)"

40303013.01 "Mandara"

40303014.04 "Gvoko (Ngoshe)"

40303015.01 "Mafa"

40303015.02 "Mofu (Moufou)"

40303015.03 "Gisiga (Guizigua)"

40303016.01 "Daba"

40303016.03 "Hina"

40303017.01 "Gudu"

40303018.02 "Gude"

40303018.03 "Nzangi (Njai)"

40303019.01 "Musgu (Mousgoum)"

40303019.02 "Gidar (Guidar)"

40303021.01 "Kotoko"

40304002.01 "Hausa (Bankal, Jerewa); Haoussa"

40304002.02 "Gwandara"

40304005.01 "Karekare"

40304006.01 "Bole (Kode, Bolawa)"

40304006.03 "Ngamo"

40304006.08 "Galambu (Galambi)"

40304006.09 "Gera (Gerawa)"

40304007.01 "Kanakuru"

40304008.01 "Tangale"

40304008.02 "Pero"

40304012.01 "Goemai (Ankwai)"

40304018.02 "Ngizim"

40304011.01 "Angas"

40304011.03 "Kofyar (Mirriam)"

40304012.02 "Montol/Mumbake"

40304014.01 "Ron (Challa, Chala)"

40304014.03 "Kulere"

40304018.01 "Bade"

40304019.02 "Pa'a (Pa awa)"

40304019.06 "Miya"

40304021.02 "Barawa (Sayawa)"

40304023.01 "Boghom"

40502018.01 "Somali"

40502019.01 "Oromo (Borana)"

40602005.00 "Arabic (Arabes choa)"

40602020.01 "Central West Gurage (Gura)"

40602004.02 "Jewish (including Hebrew)"

50201002.00 "CIRCASSIAN"

50202012.01 "Lezgi (Lesgi)"

60203001.01 "Sanskrit"

60203002.01 "Romany, Roma (gypsy)"

60203004.00 "Indian (unspecified)"

60203009.04 "Kashmiri"

60203011.01 "Marathi" 60203011.02 "Konkani"

60203012.01 "Sindhi" 60203012.02 "Lahnda (Saraiki)"

60203013.03 "Panjabi" 60203013.08 "Gujarati"

60203013.11 "Hindi" 60203013.12 "Urdu"

60203013.14 "Western Pahari (Potohari)"

60203014.01 "Nepali"

60203015.01 "Bhojpuri" 60203015.02 "Maithili"

60203016.01 "Oriya"

60203017.01 "Bengali" 60203017.02 "Assamese"

60203026.02 "Pashto (Pushto)"

60203038.01 "Talysh (Talish)"

60203040.01 "Baluchi"

60203041.00 "KURDISH"

60205000.02 "Greek"

60206007.01 "Moldovan/Rumanian"

60206012.01 "Italian"

60206016.02 "French"

60206021.01 "Spanish"

60206022.02 "Portuguese"

60208005.00 "white (unspecified)"

60208007.01 "German"

60208008.01 "Dutch" 60208008.02 "Afrikaans"

60208009.02 "English"

60209006.01 "Russian"

60209007.01 "Ukrainian"

60209014.02 "Bulgarian"

80101003.02 "Gagauz" 80101003.03 "Turkish"

80101004.01 "Azerbaijani"

80101009.04 "Kirghiz (Kyrgyz)"

110202013.01 "Telugu"

110202017.01 "Kannada"

110202022.01 "Tamil" 110202022.02 "Malayalam"

120102003.01 "Mandarin"

120202035.01 "Meithei/Manipuri"

130302056.01 "Batak"

130302103.02 "Bugis"

130302141.01 "Javanese"

130302141.02 "Sundanese"

130302143.01 "Bali"

130302149.01 "Madurese"

130302154.01 "Minangkabau"

130302156.01 "Malay"

130302156.02 "Indonesian"

170102045.04 "Chorti"

170102050.01 "Kanjobal"

170102053.02 "Mam"

170102056.02 "Kekchi (K'iche)"

170102057.01 "Quiche (Q'eqchi)"

170102057.05 "Tzutujil (Tzu'utihil)"

170102057.06 "Cakchiquel (Kaqchiquel)"

170102058.01 "Pocomam"

170102058.02 "Pocomchi (Poqomchi')"

170404000.00 "QUECHUAN"

170405000.01 "Aymara"

170502019.01 "Guarani"

-9 "other/unknown".

compute INT\_LAN6 = trunc( INT\_LANG/ 1000).

recode INT\_LAN6 (sysmis= -9).

if (INT\_LAN6 = -9) INT\_LAN6 = trunc( ETHNIC / 1000).

recode INT\_LAN6 (0 = sysmis).

val lab INT\_LAN6

10100 "KHOISAN {HADZA}"

10200 "KHOISAN {SANDAWE}"

10300 "KHOISAN {SOUTHERN AFRICA}"

10301 "KHOISAN {SOUTHERN AFRICA} NORTHERN"

10302 "KHOISAN {SOUTHERN AFRICA} CENTRAL"

10303 "KHOISAN {SOUTHERN AFRICA} SOUTHERN"

20100 "NIGER-KORDOFANIAN {KORDOFANIAN}"

20101 "NIGER-KORDOFANIAN {KORDOFANIAN} KADUGLI"

20102 "NIGER-KORDOFANIAN {KORDOFANIAN} KORDOFANIAN PROPER"

20200 "NIGER-KORDOFANIAN {NIGER-CONGO}"

20201 "NIGER-KORDOFANIAN {NIGER-CONGO} MANDE"

20202 "NIGER-KORDOFANIAN {NIGER-CONGO} NIGER-CONGO"

30100 "NILO-SAHARAN {SONGHAI}"

30200 "NILO-SAHARAN {SAHARAN}"

30201 "NILO-SAHARAN {SAHARAN} WESTERN"

30202 "NILO-SAHARAN {SAHARAN} EASTERN"

30300 "NILO-SAHARAN {MAGAN}"

30301 "NILO-SAHARAN {MAGAN} MABANG"

30400 "NILO-SAHARAN {FUR}"

30500 "NILO-SAHARAN {EAST SUDANIC}"

30501 "NILO-SAHARAN {EAST SUDANIC} EASTERN"

30502 "NILO-SAHARAN {EAST SUDANIC} WESTERN"

30503 "NILO-SAHARAN {EAST SUDANIC} NILOTIC"

30504 "NILO-SAHARAN {EAST SUDANIC} KULIAK"

30600 "NILO-SAHARAN {CENTRAL SUDANIC}"

30601 "NILO-SAHARAN {CENTRAL SUDANIC} WEST CENTRAL"

30602 "NILO-SAHARAN {CENTRAL SUDANIC} EAST CENTRAL"

30700 "NILO-SAHARAN {BERTA}"

30800 "NILO-SAHARAN {KUNAMA}"

30900 "NILO-SAHARAN {KOMUZ}"

30901 "NILO-SAHARAN {KOMUZ} GUMUZ"

30902 "NILO-SAHARAN {KOMUZ} KOMAN"

40100 "AFRO-ASIATIC {+ANCIENT EGYPTIAN}"

40200 "AFRO-ASIATIC {BERBER}"

40201 "AFRO-ASIATIC {BERBER} +GUANCHE"

40202 "AFRO-ASIATIC {BERBER} +EAST NUMIDIAN"

40203 "AFRO-ASIATIC {BERBER} BERBER PROPER"

40300 "AFRO-ASIATIC {CHADIC}"

40301 "AFRO-ASIATIC {CHADIC} MASA"

40302 "AFRO-ASIATIC {CHADIC} EAST"

40303 "AFRO-ASIATIC {CHADIC} BIU-MANDARA"

40304 "AFRO-ASIATIC {CHADIC} WEST"

40400 "AFRO-ASIATIC {OMOTIC}"

40401 "AFRO-ASIATIC {OMOTIC} SOUTH"

40402 "AFRO-ASIATIC {OMOTIC} NORTH"

40500 "AFRO-ASIATIC {CUSHITIC}"

40501 "AFRO-ASIATIC {CUSHITIC} BEJA"

40502 "AFRO-ASIATIC {CUSHITIC} CUSHITIC PROPER"

40600 "AFRO-ASIATIC {SEMITIC}"

40601 "AFRO-ASIATIC {SEMITIC} +EAST"

40602 "AFRO-ASIATIC {SEMITIC} WEST"

50100 "CAUCASIAN {SOUTH}"

50101 "CAUCASIAN {SOUTH} ZAN"

50200 "CAUCASIAN {NORTH}"

50201 "CAUCASIAN {NORTH} NORTHWEST"

50202 "CAUCASIAN {NORTH} NORTHEAST"

60100 "INDO-HITTITE {+ANATOLIAN}"

60101 "INDO-HITTITE {+ANATOLIAN} +LUWOID"

60200 "INDO-HITTITE {INDO-EUROPEAN}"

60201 "INDO-HITTITE {INDO-EUROPEAN} ARMENIAN"

60202 "INDO-HITTITE {INDO-EUROPEAN} +TOCHARIAN"

60203 "INDO-HITTITE {INDO-EUROPEAN} INDO-IRANIAN"

60204 "INDO-HITTITE {INDO-EUROPEAN} ALBANIAN"

60205 "INDO-HITTITE {INDO-EUROPEAN} GREEK"

60206 "INDO-HITTITE {INDO-EUROPEAN} ITALIC"

60207 "INDO-HITTITE {INDO-EUROPEAN} CELTIC"

60208 "INDO-HITTITE {INDO-EUROPEAN} GERMANIC"

60209 "INDO-HITTITE {INDO-EUROPEAN} BALTO-SLAVIC"

70100 "URALIC-YUKAGHIR {YUKAGHIR}"

70200 "URALIC-YUKAGHIR {URALIC}"

70201 "URALIC-YUKAGHIR {URALIC} SAMOYED"

70202 "URALIC-YUKAGHIR {URALIC} FINNO-UGRIC"

80100 "ALTAIC {ALTAIC PROPER}"

80101 "ALTAIC {ALTAIC PROPER} TURKIC"

80102 "ALTAIC {ALTAIC PROPER} MONGOLIAN-TUNGUS"

80200 "ALTAIC {KOREAN-JAPANESE}"

80201 "ALTAIC {KOREAN-JAPANESE} JAPANESE-RYUKYUAN"

90100 "CHUKCHI-KAMCHATKAN {NORTHERN}"

90101 "CHUKCHI-KAMCHATKAN {NORTHERN} CHUKCHI"

90102 "CHUKCHI-KAMCHATKAN {NORTHERN} KORYAK-ALYUTOR"

90200 "CHUKCHI-KAMCHATKAN {SOUTHERN}"

100100 "ESKIMO-ALEUT {ALEUT}"

100200 "ESKIMO-ALEUT {ESKIMO}"

100201 "ESKIMO-ALEUT {ESKIMO} INUIT"

100202 "ESKIMO-ALEUT {ESKIMO} YUPIK"

110100 "ELAMO-DRAVIDIAN {+ELAMITE}"

110200 "ELAMO-DRAVIDIAN {DRAVIDIAN}"

110201 "ELAMO-DRAVIDIAN {DRAVIDIAN} NORTHWEST"

110202 "ELAMO-DRAVIDIAN {DRAVIDIAN} DRAVIDIAN PROPER"

120100 "SINO-TIBETAN {SINITIC}"

120101 "SINO-TIBETAN {SINITIC} BAI"

120102 "SINO-TIBETAN {SINITIC} CHINESE"

120200 "SINO-TIBETAN {TIBETO-KAREN}"

120201 "SINO-TIBETAN {TIBETO-KAREN} KAREN"

120202 "SINO-TIBETAN {TIBETO-KAREN} TIBETO-BURMAN"

130100 "AUSTRIC {MIAO-YAO}"

130101 "AUSTRIC {MIAO-YAO} MIAO"

130102 "AUSTRIC {MIAO-YAO} YAO"

130200 "AUSTRIC {AUSTROASIATIC}"

130201 "AUSTRIC {AUSTROASIATIC} MUNDA"

130202 "AUSTRIC {AUSTROASIATIC} MON-KHMER"

130300 "AUSTRIC {AUSTRO-TAI}"

130301 "AUSTRIC {AUSTRO-TAI} DAIC"

130302 "AUSTRIC {AUSTRO-TAI} AUSTRONESIAN"

140100 "INDO-PACIFIC {+TASMANIAN}"

140200 "INDO-PACIFIC {ANDAMAN ISLANDS}"

140201 "INDO-PACIFIC {ANDAMAN ISLANDS} GREAT ANDAMANESE"

140202 "INDO-PACIFIC {ANDAMAN ISLANDS} SOUTH ANDAMANESE"

140300 "INDO-PACIFIC {TRANS--NEW GUINEA}"

140301 "INDO-PACIFIC {TRANS--NEW GUINEA} MAIN SECTION"

140302 "INDO-PACIFIC {TRANS--NEW GUINEA} MADANG--ADELBERT RANGE"

140303 "INDO-PACIFIC {TRANS--NEW GUINEA} TEBERAN-PAWAIAN"

140304 "INDO-PACIFIC {TRANS--NEW GUINEA} TURAMA-KIKORIAN"

140305 "INDO-PACIFIC {TRANS--NEW GUINEA} INLAND GULF"

140306 "INDO-PACIFIC {TRANS--NEW GUINEA} ELEMAN"

140307 "INDO-PACIFIC {TRANS--NEW GUINEA} TRANS-FLY--BULAKA RIVER"

140308 "INDO-PACIFIC {TRANS--NEW GUINEA} MEK"

140309 "INDO-PACIFIC {TRANS--NEW GUINEA} SENAGI"

140310 "INDO-PACIFIC {TRANS--NEW GUINEA} PAUWASI"

140311 "INDO-PACIFIC {TRANS--NEW GUINEA} NORTHERN"

140312 "INDO-PACIFIC {TRANS--NEW GUINEA} NIMBORAN"

140313 "INDO-PACIFIC {TRANS--NEW GUINEA} KAURE"

140314 "INDO-PACIFIC {TRANS--NEW GUINEA} SOUTH BIRD'S HEAD"

140315 "INDO-PACIFIC {TRANS--NEW GUINEA} KOLOPOM"

140316 "INDO-PACIFIC {TRANS--NEW GUINEA} TIMOR-ALOR-PANTAR"

140400 "INDO-PACIFIC {TRANS--NEW GUINEA} WEST PAPUAN"

140410 "INDO-PACIFIC {TRANS--NEW GUINEA} BIRD'S HEAD"

140420 "INDO-PACIFIC {TRANS--NEW GUINEA} BORAI-HATTAM"

140430 "INDO-PACIFIC {TRANS--NEW GUINEA} NORTHERN HALMAHERA"

140500 "INDO-PACIFIC {TRANS--NEW GUINEA} EAST BIRD'S HEAD"

140510 "INDO-PACIFIC {TRANS--NEW GUINEA} MEAX"

140600 "INDO-PACIFIC {TRANS--NEW GUINEA} GEELVINK BAY"

140610 "INDO-PACIFIC {TRANS--NEW GUINEA} EAST GEELVINK BAY"

140700 "INDO-PACIFIC {TRANS--NEW GUINEA} SKO"

140710 "INDO-PACIFIC {TRANS--NEW GUINEA} VANIMO"

140720 "INDO-PACIFIC {TRANS--NEW GUINEA} KRISA"

140800 "INDO-PACIFIC {TRANS--NEW GUINEA} KWOMTARI-BAIBAI"

140810 "INDO-PACIFIC {TRANS--NEW GUINEA} KWOMTARI"

140820 "INDO-PACIFIC {TRANS--NEW GUINEA} BAIBAI"

140900 "INDO-PACIFIC {TRANS--NEW GUINEA} ARAI"

141000 "INDO-PACIFIC {TRANS--NEW GUINEA} AMTO-MUSIAN"

141100 "INDO-PACIFIC {TRANS--NEW GUINEA} TORRICELLI"

141110 "INDO-PACIFIC {TRANS--NEW GUINEA} WEST WAPEI"

141120 "INDO-PACIFIC {TRANS--NEW GUINEA} WAPEI-PALEI"

141130 "INDO-PACIFIC {TRANS--NEW GUINEA} MAIMAI"

141140 "INDO-PACIFIC {TRANS--NEW GUINEA} KOMBIO-ARAPESH"

141150 "INDO-PACIFIC {TRANS--NEW GUINEA} MARIENBERG"

141160 "INDO-PACIFIC {TRANS--NEW GUINEA} MONUMBO"

141200 "INDO-PACIFIC {TRANS--NEW GUINEA} SEPIK-RAMU"

141210 "INDO-PACIFIC {TRANS--NEW GUINEA} SEPIK"

141220 "INDO-PACIFIC {TRANS--NEW GUINEA} LEONARD SCHULTZE"

141230 "INDO-PACIFIC {TRANS--NEW GUINEA} NOR-PONDO"

141240 "INDO-PACIFIC {TRANS--NEW GUINEA} RAMU"

141300 "INDO-PACIFIC {TRANS--NEW GUINEA} EAST PAPUAN"

141310 "INDO-PACIFIC {TRANS--NEW GUINEA} YELE-SOLOMONS--NEW BRITAIN"

141320 "INDO-PACIFIC {TRANS--NEW GUINEA} BOUGAINVILLE"

141330 "INDO-PACIFIC {TRANS--NEW GUINEA} REEF ISLANDS--SANTA CRUZ"

150000 "AUSTRALIAN {AUSTRALIAN}"

150100 "AUSTRALIAN {YIWAIDJAN}"

150101 "AUSTRALIAN {YIWAIDJAN} YIWAIDJIC"

150200 "AUSTRALIAN {MANGERRIAN}"

150300 "AUSTRALIAN {GUNWINYGUAN}"

150301 "AUSTRALIAN {GUNWINYGUAN} GUNWINYGIC"

150302 "AUSTRALIAN {GUNWINYGUAN} YANGMANIC"

150400 "AUSTRALIAN {BURARRAN}"

150500 "AUSTRALIAN {MARAN}"

150501 "AUSTRALIAN {MARAN} MARIC"

150600 "AUSTRALIAN {WEST BARKLY}"

150601 "AUSTRALIAN {WEST BARKLY} WAMBAYAN"

150700 "AUSTRALIAN {GARAWAN}"

150800 "AUSTRALIAN {LARAGIYAN}"

150900 "AUSTRALIAN {DALY}"

150901 "AUSTRALIAN {DALY} MOIL"

150902 "AUSTRALIAN {DALY} BRINKEN-WAGAYDY"

150903 "AUSTRALIAN {DALY} MALAKMALAK"

151000 "AUSTRALIAN {DJAMINDJUNGAN}"

151100 "AUSTRALIAN {DJERAGAN}"

151101 "AUSTRALIAN {DJERAGAN} MIRIWUNGIC"

151102 "AUSTRALIAN {DJERAGAN} KITJIC"

151200 "AUSTRALIAN {BUNABAN}"

151300 "AUSTRALIAN {NYULNYULAN}"

151400 "AUSTRALIAN {WORORAN}"

151500 "AUSTRALIAN {PAMA-NYUNGAN}"

151501 "AUSTRALIAN {PAMA-NYUNGAN} YUULNGU"

151502 "AUSTRALIAN {PAMA-NYUNGAN} TANGIC"

151503 "AUSTRALIAN {PAMA-NYUNGAN} PAMAN"

151504 "AUSTRALIAN {PAMA-NYUNGAN} YALANJIC"

151505 "AUSTRALIAN {PAMA-NYUNGAN} YIDINYIC"

151506 "AUSTRALIAN {PAMA-NYUNGAN} DYIRBALIC"

151507 "AUSTRALIAN {PAMA-NYUNGAN} NYAWAYGIC"

151508 "AUSTRALIAN {PAMA-NYUNGAN} MARIC"

151509 "AUSTRALIAN {PAMA-NYUNGAN} WAKA-KABIC"

151510 "AUSTRALIAN {PAMA-NYUNGAN} +DURUBALIC"

151511 "AUSTRALIAN {PAMA-NYUNGAN} GUMBAYNGGIRIC"

151512 "AUSTRALIAN {PAMA-NYUNGAN} YUIN-KURIC"

151513 "AUSTRALIAN {PAMA-NYUNGAN} WIRADHURIC"

151514 "AUSTRALIAN {PAMA-NYUNGAN} BAAGANDJI"

151515 "AUSTRALIAN {PAMA-NYUNGAN} +YOTAYOTIC"

151516 "AUSTRALIAN {PAMA-NYUNGAN} +KULINIC"

151517 "AUSTRALIAN {PAMA-NYUNGAN} +NGARINYERIC-YITHAYITHIC"

151518 "AUSTRALIAN {PAMA-NYUNGAN} KARNIC"

151519 "AUSTRALIAN {PAMA-NYUNGAN} WAGAYA-WARLUWARIC"

151520 "AUSTRALIAN {PAMA-NYUNGAN} KALKATUNGIC"

151521 "AUSTRALIAN {PAMA-NYUNGAN} ARANDIC"

151522 "AUSTRALIAN {PAMA-NYUNGAN} SOUTH-WEST"

160100 "NA-DENE {HAIDA}"

160200 "NA-DENE {CONTINENTAL NA-DENE}"

160201 "NA-DENE {CONTINENTAL NA-DENE} TLINGIT"

160202 "NA-DENE {CONTINENTAL NA-DENE} ATHABASKAN-EYAK"

170100 "AMERIND {NORTHERN AMERIND}"

170101 "AMERIND {NORTHERN AMERIND} ALMOSAN-KERESIOUAN"

170102 "AMERIND {NORTHERN AMERIND} PENUTIAN"

170103 "AMERIND {NORTHERN AMERIND} HOKAN"

170200 "AMERIND {CENTRAL AMERIND}"

170201 "AMERIND {CENTRAL AMERIND} TANOAN"

170202 "AMERIND {CENTRAL AMERIND} UTO-AZTECAN"

170203 "AMERIND {CENTRAL AMERIND} OTO-MANGUEAN"

170300 "AMERIND {CHIBCAN-PAEZAN}"

170301 "AMERIND {CHIBCAN-PAEZAN} CHIBCAN"

170302 "AMERIND {CHIBCAN-PAEZAN} PAEZAN"

170400 "AMERIND {ANDEAN}"

170401 "AMERIND {ANDEAN} +NORTHERN"

170402 "AMERIND {ANDEAN} URARINA-WAORANI"

170403 "AMERIND {ANDEAN} CAHUAPANAN-ZAPAROAN"

170404 "AMERIND {ANDEAN} QUECHUAN"

170405 "AMERIND {ANDEAN} AYMARAN"

170406 "AMERIND {ANDEAN} SOUTHERN"

170500 "AMERIND {EQUATORIAL-TUCANOAN}"

170501 "AMERIND {EQUATORIAL-TUCANOAN} MACRO-TUCANOAN"

170502 "AMERIND {EQUATORIAL-TUCANOAN} EQUATORIAL"

170600 "AMERIND {GE-PANO-CARIB}"

170601 "AMERIND {GE-PANO-CARIB} MACRO-CARIB"

170602 "AMERIND {GE-PANO-CARIB} GE-PANO".

compute INT\_LAN4 = trunc(INT\_LAN6/100).

val lab INT\_LAN4

101 "KHOISAN {HADZA}" 102 "KHOISAN {SANDAWE}"

103 "KHOISAN {SOUTHERN AFRICA}" 201 "NIGER-KORDOFANIAN {KORDOFANIAN}"

202 "NIGER-KORDOFANIAN {NIGER-CONGO}" 301 "NILO-SAHARAN {SONGHAI}"

302 "NILO-SAHARAN {SAHARAN}" 303 "NILO-SAHARAN {MAGAN}"

304 "NILO-SAHARAN {FUR}" 305 "NILO-SAHARAN {EAST SUDANIC}"

306 "NILO-SAHARAN {CENTRAL SUDANIC}" 307 "NILO-SAHARAN {BERTA}"

308 "NILO-SAHARAN {KUNAMA}" 309 "NILO-SAHARAN {KOMUZ}"

401 "AFRO-ASIATIC {+ANCIENT EGYPTIAN}" 402 "AFRO-ASIATIC {BERBER}"

403 "AFRO-ASIATIC {CHADIC}" 404 "AFRO-ASIATIC {OMOTIC}"

405 "AFRO-ASIATIC {CUSHITIC}" 406 "AFRO-ASIATIC {SEMITIC}"

501 "CAUCASIAN {SOUTH}" 502 "CAUCASIAN {NORTH}"

601 "INDO-HITTITE {+ANATOLIAN}" 602 "INDO-HITTITE {INDO-EUROPEAN}"

701 "URALIC-YUKAGHIR {YUKAGHIR}" 702 "URALIC-YUKAGHIR {URALIC}"

801 "ALTAIC {ALTAIC PROPER}" 802 "ALTAIC {KOREAN-JAPANESE}"

901 "CHUKCHI-KAMCHATKAN {NORTHERN}" 902 "CHUKCHI-KAMCHATKAN {SOUTHERN}"

1001 "ESKIMO-ALEUT {ALEUT}" 1002 "ESKIMO-ALEUT {ESKIMO}"

1101 "ELAMO-DRAVIDIAN {+ELAMITE}" 1102 "ELAMO-DRAVIDIAN {DRAVIDIAN}"

1201 "SINO-TIBETAN {SINITIC}" 1202 "SINO-TIBETAN {TIBETO-KAREN}"

1301 "AUSTRIC {MIAO-YAO}" 1302 "AUSTRIC {AUSTROASIATIC}"

1303 "AUSTRIC {AUSTRO-TAI}" 1401 "INDO-PACIFIC {+TASMANIAN}"

1402 "INDO-PACIFIC {ANDAMAN ISLANDS}" 1403 "INDO-PACIFIC {TRANS--NEW GUINEA}"

1500 "AUSTRALIAN {AUSTRALIAN}" 1501 "AUSTRALIAN {YIWAIDJAN}"

1502 "AUSTRALIAN {MANGERRIAN}" 1503 "AUSTRALIAN {GUNWINYGUAN}"

1504 "AUSTRALIAN {BURARRAN}" 1505 "AUSTRALIAN {MARAN}"

1506 "AUSTRALIAN {WEST BARKLY}" 1507 "AUSTRALIAN {GARAWAN}"

1508 "AUSTRALIAN {LARAGIYAN}" 1509 "AUSTRALIAN {DALY}"

1510 "AUSTRALIAN {DJAMINDJUNGAN}" 1511 "AUSTRALIAN {DJERAGAN}"

1512 "AUSTRALIAN {BUNABAN}" 1513 "AUSTRALIAN {NYULNYULAN}"

1514 "AUSTRALIAN {WORORAN}" 1515 "AUSTRALIAN {PAMA-NYUNGAN}"

1601 "NA-DENE {HAIDA}" 1602 "NA-DENE {CONTINENTAL NA-DENE}"

1701 "AMERIND {NORTHERN AMERIND}" 1702 "AMERIND {CENTRAL AMERIND}"

1703 "AMERIND {CHIBCAN-PAEZAN}" 1704 "AMERIND {ANDEAN}"

1705 "AMERIND {EQUATORIAL-TUCANOAN}" 1706 "AMERIND {GE-PANO-CARIB}".

compute INT\_LAN2 = trunc(INT\_LAN4 / 100).

val lab INT\_LAN2

1 "KHOISAN" 2 "NIGER-KORDOFANIAN"

3 "NILO-SAHARAN" 4 "AFRO-ASIATIC"

5 "CAUCASIAN" 6 "INDO-HITTITE"

7 "URALIC-YUKAGHIR" 8 "ALTAIC"

9 "CHUKCHI-KAMCHATKAN" 10 "ESKIMO-ALEUT"

11 "ELAMO-DRAVIDIAN" 12 "SINO-TIBETAN"

13 "AUSTRIC, AUSTROASIATIC, AUSTRO-TAI" 14 "INDO-PACIFIC"

15 "AUSTRALIAN" 16 "NA-DENE"

17 "AMERIND".

var lab INT\_LANG 'Language used in interview'.

var lab INT\_LAN6 'Language group of respondent - 6 digit'.

var lab INT\_LAN4 'Language group of respondent - 4 digit'.

var lab INT\_LAN2 'Language group of respondent - 2 digit'.

\* Below is an alternative system, from ‘World Values Survey’ website; it is less detailed than the Ruhlen system.

\*var lab ETHNIC\_gp 'Ethnic group of respondent'.

\* The following set of value labels are from 'World Values Survey', file 'XWVSEVS\_1981\_2000\_v20060423.sav'

\* variable x051 - I added labels in "double quotes".

value label ETHNIC\_gp

10 'Afghanistan'

20 'Albanian'

30 'Armenian'

40 'Asian'

50 'Asian - Central (Saudi)'

60 'Asian - Central (Yemeni)'

70 'Asian - Central (Arabic)'

80 'Asian - East (Chinese, Japanese)'

90 'Asian - South (Indian, Hindu, Pakistani, Bangladeshi)'

100 'Australian - Aboriginal '

110 'Avarian'

120 'Azerbaijanian'

130 'Bangladeshi'

140 'Bangladeshi-Baluchi'

150 'Bangladeshi-Pathan'

160 'Bangladeshi-Punjabi'

170 'Bangladeshi-Sindhi'

180 'Bangladeshi-Urdu Speaking'

190 'Belorussian'

200 'Black African'

210 'Black-Caribbean'

220 'Black-Other / Black'

230 'Bulgarian'

240 'China - Han nationality'

250 'China - Hui nationality'

260 'China - Manchu nationality'

270 'China - Miao nationality'

280 'China - Uygur nationality'

290 'China - Zhuang nationality'

299 "Chinese"

300 'Christian'

310 'Coloured (medium)'

320 'Coloured (dark)'

330 'Coloured (light)'

340 'Croatian'

350 'Crosbreed; Mulato'

360 'Czech'

370 'Egyptian - From Nile Delta'

380 'Egyptian - From Upper Egypt'

390 'Egyptian - Noba'

400 'Egyptian/Sudanese'

410 'European'

420 'European - English'

430 'European - South'

440 'European - Swiss Italian'

450 'European - Swiss French'

460 'European - Swiss German'

470 'French'

480 'Gagaus'

490 'Georgian'

500 'German'

510 'Greek'

520 'Gypsy'

522 'Half breed of black and white'

524 'Half breed of white and indian'

530 'Hungarian'

540 'Indian - Buddhists'

550 'Indian - Dominant Peasant Castes'

560 'Indian - Peasant Backward Castes'

570 'Indian - Scheduled Castes'

580 'Indian - Scheduled Tribus'

590 'Indian - Service Backward Castes'

600 'Indian - Sikhs'

610 'Indian - Upper Castes'

620 'Indian - Weavers and Craftsman'

623 'Indian - No caste'

625 'Indian'

630 'Indian (American)'

640 'Indigenous'

650 'Iranian'

660 'Italian'

670 'Javanese'

680 'Jew'

690 'Jordan'

700 'Jordanian/Palestinian'

710 'Keldan'

720 'Kurd/Esid'

730 'Latvian'

740 'Lebanese/Syrian'

750 'Lezgin'

760 'Lithuanian'

770 'Macedonian'

780 'Malay'

790 'Moldovian'

800 'Muslim'

810 'Pakistani'

820 'Palestinan'

830 'Peruan - Crossbreed Amazones'

840 'Peruan - Crossbreed Andino'

850 'Philippines'

860 'Philippines-Aklanon'

870 'Philippines-Antiqueno'

880 'Philippines-Bagubu/Bagobo'

890 'Philippines-Bicolano'

900 'Philippines-Bilaan'

910 'Philippines-Bisaya'

920 'Philippines-Boholano'

930 'Philippines-Bungolanon'

940 'Philippines-Cebuano'

950 'Philippines-Chabacano'

960 'Philippines-Davaoeno/Dabawenyo'

970 'Philippines-Ibaloy/Ibanag'

980 'Philippines-Ilocano'

990 'Philippines-Ilonggo'

1000 'Philippines-Kalagan'

1010 'Philippines-Kanglo'

1020 'Philippines-Kankana-ay'

1030 'Philippines-Kapampangan'

1040 'Philippines-Kaulo'

1050 'Philippines-Kulanan'

1060 'Philippines-Litinya/Leyteyo'

1070 'Philippines-Lubano'

1080 'Philippines-Maguindanao'

1090 'Philippines-Manobo'

1100 'Philippines-Maranao/Matanao'

1110 'Philippines-Masbateno'

1120 'Philippines-Minority'

1130 'Philippines-Pangasinense'

1140 'Philippines-Sama'

1150 'Philippines-Suriganon'

1160 'Philippines-Tagalog'

1170 'Philippines-Tausog'

1180 'Philippines-Waray'

1190 'Philippines-Zambal'

1200 'Polish'

1210 'Romanian'

1220 'Russian'

1230 'Serbian'

1240 'Slovenian'

1250 'Spanish; Hispanic (& Portuguese)'

1260 'Sumatranese'

1270 'Sundanese'

1280 'Tatarian'

1290 'Turkish'

1300 'Turkmenian'

1310 'Ukrainian'

1320 'Vietnam - Dao'

1330 'Vietnam - Ede'

1340 'Vietnam - Hoa'

1350 'Vietnam - Kinh'

1360 'Vietnam - Mong'

1370 'Vietnam - Muong'

1380 'Vietnam - Ray'

1390 'Vlav'

1400 'White / Caucasian White'

1420 'Yugoslavian'

8000 'other'

-9 'missing'.

\* The label for 625 'Indian - No religion, No caste' is used in WVS; I assume this means Indians in general.

\* WVS use these codes: -5 'Missing; Unknown' -4 'Not asked' -3 'Not applicable' -2 'No answer' -1 "Don´t know".