# Appendix A: Language sample

Information on the language sample used in the study is listed in Tables A1-A4.

**Key to reading table:**

**ISO 693-3:** ISO 693-3 code for language used in survey.

**Language:** Dialect is given in parentheses where relevant.

**Syllable Structure:**

***S*** = Simple

***MC*** = Moderately Complex

***C*** = Complex

***HC*** = Highly Complex

**Macro-area:** Following Dryer (1989: 268; 1992: 83, 133-5).

***Africa =*** continent of Africa, including Semitic languages of southwest Asia.

***Australia & New Guinea =*** Australian continent and Melanesia, excluding Austronesian languages of Melanesia.

***Eurasia =*** Eurasian landmass, excluding Semitic and languages from families of southeast Asia as defined below, and including the Munda languages of Austro-Asiatic.

***North America =*** North American continent, including languages of Mexico, Mayan and Aztecan languages in Central America, and some branches of Chibchan-Paezan.

***South America =*** South American continent, including languages of Central America except Mayan and Aztecan languages, and some Chibchan-Paezan branches.

***Southeast Asia & Oceania =*** Southeast Asian region, including all Sino-Tibetan, Tai-Kadai, Hmong-Mien, and Austro-Asiatic languages excluding Munda, and Oceania region (Austronesian languages).

**Top-level family** and **Subfamily:** Following genealogical classifications listed in Glottolog 3.3 (Hammarström et al. 2018).

**Speaker Population:** L1 speaker population figure for language (or specific dialect) given in Ethnologue 21 (Simons & Fennig 2018). An asterisk indicates that another source was used for population estimate; these can be found beneath the table.

**Date:** Date given in Ethnologue 21 (Simons & Fennig 2018) for speaker population figure.

**Vitality Status:** Following Ethnologue 21 (Simons & Fennig 2018).

***Institutional =*** language has wide use in the home and community and official status at educational, provincial, national, and/or international levels.

***Developing =*** language is used in the home, community, and sometimes broader contexts, and in initial stages of developing a system of writing and standardization.

***Vigorous =*** language is used in the home and community by speakers of all generations, but has not yet developed a system of graphization or standardization*.*

***In Trouble =*** language is currently in the process of losing intergenerational transmission, with the community shifting to other languages for daily use, but there are still speakers of child-bearing age.

***Dying =*** language has lost intergenerational transmission entirely, and all fluent speakers are above child-bearing age.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ISO 639-3** | **Language** | **Syllable Structure** | **Macro-area** | **Top-level family** | **Subfamily** | **Speaker Population** | **Date** | **Vitality**  **Status** |
| hts | **Hadza** | S | Africa | (isolate) |  | 950 | 2013 | In Trouble |
| grj | **Southern Grebo** | S | Africa | Atlantic-Congo | *Volta-Congo* | 65,000 | 2012 | Vigorous |
| yor | **Yoruba** | S | Africa | Atlantic-Congo | *Volta-Congo* | 19,043,700 | 1993 | Institutional |
| mhi | **Ma’di** | S | Africa | Central Sudanic | *Moru-Madi* | 293,000 | 2014 | Developing |
| bbo | **Southern Bobo Madaré** | S | Africa | Mande | *Western Mande* | 181,000 | 2009 | Developing |
| svs | **Savosavo** | S | Aus & New Guinea | (isolate) |  | 2,420 | 1999 | Vigorous |
| kbk | **Grass Koiari** | S | Aus & New Guinea | Koiarian | *Koiaric* | 1,700 | 2000 | Vigorous |
| roo | **Rotokas** | S | Aus & New Guinea | North Bougainville | *Rotokas-Askopan* | 4,320 | 1981 | Developing |
| kjs | **East Kewa** | S | Aus & New Guinea | Nuclear Trans New Guinea | *Enga-Kewa-Huli* | 45,000 | 2000 | Developing |
| tow | **Towa** | S | N America | Kiowa-Tanoan |  | 1,790 | 2007 | In Trouble |
| mio | **Pinotepa Mixtec** | S | N America | Otomanguean | *Eastern Otomanguean* | 20,000 | 1990 | Vigorous |
| ute | **Ute** | S | N America | Uto-Aztecan | *Northern Uto-Aztecan* | 920 | 2007 | In Trouble |
| ura | **Urarina** | S | S America | (isolate) |  | 3,000 | 2002 | Developing |
| wba | **Warao** | S | S America | (isolate) |  | 28,100 | 2007 | Vigorous |
| apu | **Apurinã** | S | S America | Arawakan | *Southern Maipuran* | 2,870 | 2006 | In Trouble |
| huu | **Murui Huitoto** | S | S America | Huitotoan | *Nuclear Witotoan* | 2,000 | 2016 | In Trouble |
| cav | **Cavineña** | S | S America | Pano-Tacanan | *Tacanan* | 600 | 2011 | In Trouble |
| cub | **Cubeo** | S | S America | Tucanoan | *Eastern Tucanoan* | 6,260 | 2008 | Institutional |
| dru | **Rukai (Budai dialect)** | S | SE Asia & Oceania | Austronesian |  | 10,500 | 2002 | Developing |
| mri | **Maori** | S | SE Asia & Oceania | Austronesian | *Malayo-Polynesian* | 158,640 | 2013 | In Trouble |
| khc | **Tukang Besi North** | S | SE Asia & Oceania | Austronesian | *Malayo-Polynesian* | 120,000 | 1995 | Vigorous |
| sxr | **Saaroa** | S | SE Asia & Oceania | Austronesian | *Tsouic* | 10 | 2012 | Dying |
| iii | **Sichuan Yi** | S | SE Asia & Oceania | Sino-Tibetan | *Burmo-Qiangic* | 2,000,000 | 2004 | Institutional |

**Table A1.** Portion of language sample with Simple syllable structure.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ISO 639-3** | **Language** | **Syllable Structure** | **Macro-area** | **Top-level family** | **Subfamily** | **Speaker Population** | **Date** | **Vitality**  **Status** |
| ktb | **Kambaata** | MC | Africa | Afro-Asiatic | *Cushitic* | 743,000 | 2007 | Institutional |
| ewe | **Ewe** | MC | Africa | Atlantic-Congo | *Volta-Congo* | 4,184,000 | 2013 | Institutional |
| fvr | **Fur** | MC | Africa | Furan |  | 745,800 | 2004 | Developing |
| knc | **Kanuri** | MC | Africa | Saharan | *Western Saharan* | 3,290,500 | 1985 | Institutional |
| ayz | **Maybrat** | MC | Aus & New Guinea | Maybrat-Karon |  | 20,000 | 1987 | Developing |
| kms | **Kamasau** | MC | Aus & New Guinea | Nuclear Torricelli | *Marienberg* | 960 | 2003 | In Trouble |
| spl | **Selepet** | MC | Aus & New Guinea | Nuclear Trans New Guinea | *Finisterre-Huon* | 7,000 | 1988 | Developing |
| aly | **Alyawarra** | MC | Aus & New Guinea | Pama-Nyungan | *Arandic-Thura-Yura* | 1,660 | 2006 | Developing |
| khr | **Kharia** | MC | Eurasia | Austroasiatic | *Mundaic* | 241,580 | 2001 | Developing |
| tel | **Telugu** | MC | Eurasia | Dravidian | *South Dravidian* | 74,244,300 | 2001 | Institutional |
| dry | **Darai** | MC | Eurasia | Indo-European | *Indo-Iranian* | 11,700 | 2011 | In Trouble |
| mjg | **Tu** | MC | Eurasia | Mongolic | *Southern Periphery Mongolic* | 152,000 | 2000 | In Trouble |
| kca | **Eastern Khanty** | MC | Eurasia | Uralic | *Khantyic* | 2,000 | 2007 | In Trouble |
| kyh | **Karok** | MC | N America | (isolate) |  | 12 | 2007 | Dying |
| scs | **North Slavey (Hare dialect)** | MC | N America | Athabaskan-Eyak-Tlingit | *Athabaskan-Eyak* | 710 | 2007 | In Trouble |
| kal | **Kalaallisut** | MC | N America | Eskimo-Aleut | *Eskimo* | 44,000 | 2007 | Institutional |
| cho | **Choctaw** | MC | N America | Muskogean | *Western Muskogean* | 10,400 | 2010 | In Trouble |
| car | **Carib** | MC | S America | Cariban | *Guianan* | 7,358 | 2001 | In Trouble |
| qvi | **Imbabura Highland Quechua** | MC | S America | Quechuan | *Quechua II* | 150,000 | 2007 | Developing |
| cod | **Cocama-Cocamilla** | MC | S America | Tupian | *Maweti-Guarani* | 250 | 2007 | Dying |
| pac | **Pacoh** | MC | SE Asia & Oceania | Austroasiatic | *Katuic* | 32,500 | 2002 | In Trouble |
| pwn | **Paiwan** | MC | SE Asia & Oceania | Austronesian |  | 66,100 | 2002 | Developing |
| mji | **Kim Mun (Vietnam dialect)** | MC | SE Asia & Oceania | Hmong-Mien | *Mienic* | 374,500 | 2000 | Vigorous |
| aot | **Atong** | MC | SE Asia & Oceania | Sino-Tibetan | *Brahmaputran* | 10,000 | (no date) | In Trouble |
| yue | **Cantonese** | MC | SE Asia & Oceania | Sino-Tibetan | *Sinitic* | 62,967,910 | 2013 | Institutional |
| lao | **Lao** | MC | SE Asia & Oceania | Tai-Kadai | *Kam-Tai* | 3,253,700 | 2005 | Institutional |

**Table A2.** Portion of language sample with Moderately Complex syllable structure.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ISO 639-3** | **Language** | **Syllable Structure** | **Macro-area** | **Top-level family** | **Subfamily** | **Speaker Population** | **Date** | **Vitality**  **Status** |
| mpi | **Mpade (Makari dialect)** | C | Africa | Afro-Asiatic | *Chadic* | 16,000 | 2004 | In Trouble |
| dyo | **Jola-Fonyi** | C | Africa | Atlantic-Congo | *North-Central Atlantic* | 397,100 | (no date) | Developing |
| lun | **Lunda** | C | Africa | Atlantic-Congo | *Volta-Congo* | 403,000 | 2010 | Institutional |
| mdx | **Dizin (Central dialect)** | C | Africa | Dizoid |  | 33,900 | 2010 | Institutional |
| tbi | **Gaam** | C | Africa | Eastern Jebel |  | 67,200 | 2000 | Vigorous |
| mpc | **Mangarrayi** | C | Aus & New Guinea | Mangarrayi-Maran |  | 12 | 2006 | Dying |
| nir | **Nimboran** | C | Aus & New Guinea | Nimboranic |  | 2,000 | 1987 | Dying |
| opm | **Oksapmin** | C | Aus & New Guinea | Nuclear Trans New Guinea | *Asman-Awyu-Ok* | 8,000 | 1991 | Developing |
| bcj | **Bardi** | C | Aus & New Guinea | Nyulnyulan | *Western Nyulnyulan* | 160 | 2006 | Dying |
| ung | **Ngarinyin** | C | Aus & New Guinea | Worrorran |  | 57 | 2006 | In Trouble |
| bsk | **Burushaski** | C | Eurasia | (isolate) |  | 96,800 | 2004 | Vigorous |
| eus | **Basque** | C | Eurasia | (isolate) |  | 545,800 | 2012 | Institutional |
| niv | **Nivkh (West Sakhalin dialect)** | C | Eurasia | (isolate) |  | 15\* | 2014 | Dying |
| bak | **Bashkir** | C | Eurasia | Turkic | *Common Turkic* | 1,245,990 | 2010 | Institutional |
| ket | **Ket** | C | Eurasia | Yeniseian | *Northern Yeniseian* | 2010 | 2010 | Dying |
| pay | **Pech** | C | N America | Chibchan |  | 990 | 1993 | Dying |
| tzh | **Tzeltal** | C | N America | Mayan | *Core Mayan* | 372,000 | 2000 | Developing |
| lkt | **Lakota** | C | N America | Siouan | *Core Siouan* | 2,200 | 1997 | In Trouble |
| kbc | **Kadiwéu** | C | S America | Guaicuruan |  | 1,590 | 2006 | In Trouble |
| wmd | **Mamaindê** | C | S America | Nambiquaran | *Nambikwara Complex* | 330 | 2007 | In Trouble |
| apn | **Apinayé** | C | S America | Nuclear-Macro-Je | *Je* | 1,260 | 2003 | Developing |
| cap | **Chipaya** | C | S America | Uru-Chipaya |  | 1,200 | 1995 | Developing |
| kpm | **Koho** | C | SE Asia & Oceania | Austroasiatic | *Bahnaric* | 166,000 | 2009 | Developing |
| lpa | **Lelepa** | C | SE Asia & Oceania | Austronesian | *Malayo-Polynesian* | 400 | 1989 | Vigorous |
| lep | **Lepcha** | C | SE Asia & Oceania | Sino-Tibetan | *Himalayish* | 69,800 | 2001 | Vigorous |

**Table A3.** Portion of language sample with Complex syllable structure.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ISO 639-3** | **Language** | **Syllable Structure** | **Macro-area** | **Top-level family** | **Subfamily** | **Speaker Population** | **Date** | **Vitality**  **Status** |
| shi | **Tashlhiyt** | HC | Africa | Afro-Asiatic | *Berber* | 3,896,000 | 2004 | Developing |
| dow | **Doyayo** | HC | Africa | Atlantic-Congo | *Volta-Congo* | 18,000 | 1985 | Developing |
| bcq | **Bench** | HC | Africa | Ta-Ne-Omotic |  | 348,000 | 2007 | Institutional |
| mcr | **Menya** | HC | Aus & New Guinea | Angan | *Nuclear Angan* | 20,000 | 1998 | Developing |
| kjn | **Kunjen** | HC | Aus & New Guinea | Pama-Nyungan | *Paman* | 20 | 1991 | Dying |
| amp | **Alamblak** | HC | Aus & New Guinea | Sepik | *Sepik Hill* | 1,530 | 2000 | Developing |
| wut | **Wutung** | HC | Aus & New Guinea | Sko | *Nuclear Skou-Serra-Piore* | 900 | 2003 | Vigorous |
| kbd | **Kabardian** | HC | Eurasia | Abkhaz-Adyge | *Circassian* | 1,628,500 | 2010 | Developing |
| itl | **Itelmen** | HC | Eurasia | Chukotko-Kamchatkan |  | 80 | 2010 | Dying |
| als | **Albanian (Tosk dialect)** | HC | Eurasia | Indo-European | *Albanian* | 1,841,400 | 2012 | Institutional |
| pol | **Polish** | HC | Eurasia | Indo-European | *Balto-Slavic* | 40,248,740 | 2013 | Institutional |
| kat | **Georgian** | HC | Eurasia | Kartvelian | *Georgian-Zan* | 4,347,320 | 1993 | Institutional |
| lez | **Lezgian** | HC | Eurasia | Nakh-Daghestanian | *Daghestanian* | 616,760 | 2010 | Institutional |
| pqm | **Passamaquoddy-Maliseet** | HC | N America | Algic | *Algonquian* | 590 | 2011 | In Trouble |
| coc | **Cocopa** | HC | N America | Cochimi-Yuman | *Yuman* | 350 | 1998 | In Trouble |
| moh | **Mohawk** | HC | N America | Iroquoian | *Northern Iroquoian* | 3,540 | 1999 | In Trouble |
| yak | **Yakama Sahaptin** | HC | N America | Sahaptian | *Sahaptin* | 5\*\* | 2006 | Dying |
| thp | **Thompson** | HC | N America | Salishan | *Interior Salish* | 130 | 2014 | In Trouble |
| ood | **Tohono O’odham** | HC | N America | Uto-Aztecan | *Southern Uto-Aztecan* | 14,094 | 2007 | In Toruble |
| nuk | **Nuu-chah-nulth** | HC | N America | Wakashan | *Southern Wakashan* | 130 | 2014 | Dying |
| kbh | **Camsá** | HC | S America | (isolate) |  | 4,000 | 2008 | Developing |
| pib | **Yine** | HC | S America | Arawakan | *Southern Maipuran* | 4,000 | 2000 | Developing |
| teh | **Tehuelche** | HC | S America | Chonan | *Continental Chonan* | 5\*\*\* | 2012 | Dying |
| alc | **Qawasqar** | HC | S America | Kawesqar | *North Central Alacalufan* | 12 | 2006 | Dying |
| sea | **Semai** | HC | SE Asia & Oceania | Austroasiatic | *Aslian* | 10,000 | 2007 | Institutional |

**Table A4.** Portion of language sample with Highly Complex syllable structure.

\* Population figure from Botma & Shiraishi (2014).

\*\* Population figure from Hargus & Beavert (2006).

\*\*\* Population figure from *aoNEK FILMS* (2012), includes semi-speakers.