

## 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 Aztec languages in Central America, and some branches of Chibchan-Paezan.

*South America* = South American continent, including languages of Central America except Mayan and Aztec 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	<b>Hadza</b>	S	Africa	(isolate)		950	2013	In Trouble
grj	<b>Southern Grebo</b>	S	Africa	Atlantic-Congo	<i>Volta-Congo</i>	65,000	2012	Vigorous
yor	<b>Yoruba</b>	S	Africa	Atlantic-Congo	<i>Volta-Congo</i>	19,043,700	1993	Institutional
mhi	<b>Ma'di</b>	S	Africa	Central Sudanic	<i>Moru-Madi</i>	293,000	2014	Developing
bbo	<b>Southern Bobo Madaré</b>	S	Africa	Mande	<i>Western Mande</i>	181,000	2009	Developing
svs	<b>Savosavo</b>	S	Aus & New Guinea	(isolate)		2,420	1999	Vigorous
kbk	<b>Grass Koiari</b>	S	Aus & New Guinea	Koian	<i>Koian</i>	1,700	2000	Vigorous
roo	<b>Rotokas</b>	S	Aus & New Guinea	North Bougainville	<i>Rotokas-Askopan</i>	4,320	1981	Developing
kjs	<b>East Kewa</b>	S	Aus & New Guinea	Nuclear Trans New Guinea	<i>Enga-Kewa-Huli</i>	45,000	2000	Developing
tow	<b>Towa</b>	S	N America	Kiowa-Tanoan		1,790	2007	In Trouble
mio	<b>Pinotepa Mixtec</b>	S	N America	Otomanguean	<i>Eastern Otomanguean</i>	20,000	1990	Vigorous
ute	<b>Ute</b>	S	N America	Uto-Aztecan	<i>Northern Uto-Aztecan</i>	920	2007	In Trouble
ura	<b>Urarina</b>	S	S America	(isolate)		3,000	2002	Developing
wba	<b>Warao</b>	S	S America	(isolate)		28,100	2007	Vigorous
apu	<b>Apurinã</b>	S	S America	Arawakan	<i>Southern Maipuran</i>	2,870	2006	In Trouble
huu	<b>Murui Huitoto</b>	S	S America	Huitotoan	<i>Nuclear Witotoan</i>	2,000	2016	In Trouble
cav	<b>Cavineña</b>	S	S America	Pano-Tacanan	<i>Tacanan</i>	600	2011	In Trouble
cub	<b>Cubeo</b>	S	S America	Tucanoan	<i>Eastern Tucanoan</i>	6,260	2008	Institutional
dru	<b>Rukai (Budai dialect)</b>	S	SE Asia & Oceania	Austronesian		10,500	2002	Developing
mri	<b>Maori</b>	S	SE Asia & Oceania	Austronesian	<i>Malayo-Polynesian</i>	158,640	2013	In Trouble
khc	<b>Tukang Besi North</b>	S	SE Asia & Oceania	Austronesian	<i>Malayo-Polynesian</i>	120,000	1995	Vigorous
sxr	<b>Saaroa</b>	S	SE Asia & Oceania	Austronesian	<i>Tsouic</i>	10	2012	Dying
iii	<b>Sichuan Yi</b>	S	SE Asia & Oceania	Sino-Tibetan	<i>Burmo-Qiangic</i>	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	<b>Kambaata</b>	MC	Africa	Afro-Asiatic	<i>Cushitic</i>	743,000	2007	Institutional
ewe	<b>Ewe</b>	MC	Africa	Atlantic-Congo	<i>Volta-Congo</i>	4,184,000	2013	Institutional
fvr	<b>Fur</b>	MC	Africa	Furan		745,800	2004	Developing
knc	<b>Kanuri</b>	MC	Africa	Saharan	<i>Western Saharan</i>	3,290,500	1985	Institutional
ayz	<b>Maybrat</b>	MC	Aus & New Guinea	Maybrat-Karon		20,000	1987	Developing
kms	<b>Kamasau</b>	MC	Aus & New Guinea	Nuclear Torricelli	<i>Marienberg</i>	960	2003	In Trouble
spl	<b>Selepet</b>	MC	Aus & New Guinea	Nuclear Trans New Guinea	<i>Finisterre-Huon</i>	7,000	1988	Developing
aly	<b>Alyawarra</b>	MC	Aus & New Guinea	Pama-Nyungan	<i>Arandic-Thura-Yura</i>	1,660	2006	Developing
kh	<b>Kharia</b>	MC	Eurasia	Austroasiatic	<i>Mundaic</i>	241,580	2001	Developing
tel	<b>Telugu</b>	MC	Eurasia	Dravidian	<i>South Dravidian</i>	74,244,300	2001	Institutional
dry	<b>Darai</b>	MC	Eurasia	Indo-European	<i>Indo-Iranian</i>	11,700	2011	In Trouble
mjg	<b>Tu</b>	MC	Eurasia	Mongolic	<i>Southern Periphery Mongolic</i>	152,000	2000	In Trouble
kca	<b>Eastern Khanty</b>	MC	Eurasia	Uralic	<i>Khantyic</i>	2,000	2007	In Trouble
kyh	<b>Karok</b>	MC	N America	(isolate)		12	2007	Dying
scs	<b>North Slavey (Hare dialect)</b>	MC	N America	Athabaskan-Eyak-Tlingit	<i>Athabaskan-Eyak</i>	710	2007	In Trouble
kal	<b>Kalaallisut</b>	MC	N America	Eskimo-Aleut	<i>Eskimo</i>	44,000	2007	Institutional
cho	<b>Choctaw</b>	MC	N America	Muskogean	<i>Western Muskogean</i>	10,400	2010	In Trouble
car	<b>Carib</b>	MC	S America	Cariban	<i>Guianan</i>	7,358	2001	In Trouble
qvi	<b>Imbabura Highland Quechua</b>	MC	S America	Quechuan	<i>Quechua II</i>	150,000	2007	Developing
cod	<b>Cocama-Cocamilla</b>	MC	S America	Tupian	<i>Maweti-Guarani</i>	250	2007	Dying
pac	<b>Pacoh</b>	MC	SE Asia & Oceania	Austroasiatic	<i>Katuic</i>	32,500	2002	In Trouble
pwn	<b>Paiwan</b>	MC	SE Asia & Oceania	Austronesian		66,100	2002	Developing
mji	<b>Kim Mun (Vietnam dialect)</b>	MC	SE Asia & Oceania	Hmong-Mien	<i>Mienic</i>	374,500	2000	Vigorous
aot	<b>Atong</b>	MC	SE Asia & Oceania	Sino-Tibetan	<i>Brahmaputran</i>	10,000	(no date)	In Trouble
yue	<b>Cantonese</b>	MC	SE Asia & Oceania	Sino-Tibetan	<i>Sinitic</i>	62,967,910	2013	Institutional
lao	<b>Lao</b>	MC	SE Asia & Oceania	Tai-Kadai	<i>Kam-Tai</i>	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	<b>Mpade (Makari dialect)</b>	C	Africa	Afro-Asiatic	<i>Chadic</i>	16,000	2004	In Trouble
dyo	<b>Jola-Fonyi</b>	C	Africa	Atlantic-Congo	<i>North-Central Atlantic</i>	397,100	(no date)	Developing
lun	<b>Lunda</b>	C	Africa	Atlantic-Congo	<i>Volta-Congo</i>	403,000	2010	Institutional
mdx	<b>Dizin (Central dialect)</b>	C	Africa	Dizoid		33,900	2010	Institutional
tbi	<b>Gaam</b>	C	Africa	Eastern Jebel		67,200	2000	Vigorous
mpc	<b>Mangarrayi</b>	C	Aus & New Guinea	Mangarrayi-Maran		12	2006	Dying
nir	<b>Nimboran</b>	C	Aus & New Guinea	Nimboranic		2,000	1987	Dying
opm	<b>Oksapmin</b>	C	Aus & New Guinea	Nuclear Trans New Guinea	<i>Asman-Awyu-Ok</i>	8,000	1991	Developing
bcj	<b>Bardi</b>	C	Aus & New Guinea	Nyulnyulan	<i>Western Nyulnyulan</i>	160	2006	Dying
ung	<b>Ngarinyin</b>	C	Aus & New Guinea	Worrorran		57	2006	In Trouble
bsk	<b>Burushaski</b>	C	Eurasia	(isolate)		96,800	2004	Vigorous
eus	<b>Basque</b>	C	Eurasia	(isolate)		545,800	2012	Institutional
niv	<b>Nivkh (West Sakhalin dialect)</b>	C	Eurasia	(isolate)		15*	2014	Dying
bak	<b>Bashkir</b>	C	Eurasia	Turkic	<i>Common Turkic</i>	1,245,990	2010	Institutional
ket	<b>Ket</b>	C	Eurasia	Yeniseian	<i>Northern Yeniseian</i>	2010	2010	Dying
pay	<b>Pech</b>	C	N America	Chibchan		990	1993	Dying
tzh	<b>Tzeltal</b>	C	N America	Mayan	<i>Core Mayan</i>	372,000	2000	Developing
lkt	<b>Lakota</b>	C	N America	Siouan	<i>Core Siouan</i>	2,200	1997	In Trouble
kbc	<b>Kadiwéu</b>	C	S America	Guaicuruan		1,590	2006	In Trouble
wmd	<b>Mamaindê</b>	C	S America	Nambiquaran	<i>Nambikwara Complex</i>	330	2007	In Trouble
apn	<b>Apinayé</b>	C	S America	Nuclear-Macro-Je	<i>Je</i>	1,260	2003	Developing
cap	<b>Chipaya</b>	C	S America	Uru-Chipaya		1,200	1995	Developing
kpm	<b>Koho</b>	C	SE Asia & Oceania	Austroasiatic	<i>Bahnaric</i>	166,000	2009	Developing
lpa	<b>Lelepa</b>	C	SE Asia & Oceania	Austronesian	<i>Malayo-Polynesian</i>	400	1989	Vigorous
lep	<b>Lepcha</b>	C	SE Asia & Oceania	Sino-Tibetan	<i>Himalayish</i>	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	<b>Tashlhiyt</b>	HC	Africa	Afro-Asiatic	<i>Berber</i>	3,896,000	2004	Developing
dow	<b>Doyayo</b>	HC	Africa	Atlantic-Congo	<i>Volta-Congo</i>	18,000	1985	Developing
bcq	<b>Bench</b>	HC	Africa	Ta-Ne-Omoti		348,000	2007	Institutional
mcr	<b>Menya</b>	HC	Aus & New Guinea	Angan	<i>Nuclear Angan</i>	20,000	1998	Developing
kjn	<b>Kunjen</b>	HC	Aus & New Guinea	Pama-Nyungan	<i>Paman</i>	20	1991	Dying
amp	<b>Alamblak</b>	HC	Aus & New Guinea	Sepik	<i>Sepik Hill</i>	1,530	2000	Developing
wut	<b>Wutung</b>	HC	Aus & New Guinea	Sko	<i>Nuclear Skou-Serra-Piore</i>	900	2003	Vigorous
kbd	<b>Kabardian</b>	HC	Eurasia	Abkhaz-Adyghe	<i>Circassian</i>	1,628,500	2010	Developing
itl	<b>Itelmen</b>	HC	Eurasia	Chukotko-Kamchatkan		80	2010	Dying
als	<b>Albanian (Tosk dialect)</b>	HC	Eurasia	Indo-European	<i>Albanian</i>	1,841,400	2012	Institutional
pol	<b>Polish</b>	HC	Eurasia	Indo-European	<i>Balto-Slavic</i>	40,248,740	2013	Institutional
kat	<b>Georgian</b>	HC	Eurasia	Kartvelian	<i>Georgian-Zan</i>	4,347,320	1993	Institutional
lez	<b>Lezgian</b>	HC	Eurasia	Nakh-Daghestanian	<i>Daghestanian</i>	616,760	2010	Institutional
pqm	<b>Passamaquoddy-Maliseet</b>	HC	N America	Algonquian	<i>Algonquian</i>	590	2011	In Trouble
coc	<b>Cocopa</b>	HC	N America	Cochimi-Yuman	<i>Yuman</i>	350	1998	In Trouble
moh	<b>Mohawk</b>	HC	N America	Iroquoian	<i>Northern Iroquoian</i>	3,540	1999	In Trouble
yak	<b>Yakama Sahaptin</b>	HC	N America	Sahaptian	<i>Sahaptin</i>	5**	2006	Dying
thp	<b>Thompson</b>	HC	N America	Salishan	<i>Interior Salish</i>	130	2014	In Trouble
ood	<b>Tohono O'odham</b>	HC	N America	Uto-Aztecan	<i>Southern Uto-Aztecan</i>	14,094	2007	In Trouble
nuk	<b>Nuu-chah-nulth</b>	HC	N America	Wakashan	<i>Southern Wakashan</i>	130	2014	Dying
kbh	<b>Camsá</b>	HC	S America	(isolate)		4,000	2008	Developing
pib	<b>Yine</b>	HC	S America	Arawakan	<i>Southern Maipuran</i>	4,000	2000	Developing
teh	<b>Tehuelche</b>	HC	S America	Chonan	<i>Continental Chonan</i>	5***	2012	Dying
alc	<b>Qawasqar</b>	HC	S America	Kawesqar	<i>North Central Alacalufan</i>	12	2006	Dying
sea	<b>Semai</b>	HC	SE Asia & Oceania	Austroasiatic	<i>Aslian</i>	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.