Relation between Word Order of Languages and the Entropy of Y-chromosome Haplogroup Distribution of the Speakers' Population

言語の語順とその話者集団のY染色体ハプログループ分布のエントロピーとの関係

Terumasa EHARA
江原暉将

Ehara NLP Research Laboratory
江原自然言語処理研究室
http://www.ne.jp/asahi/eharate/eharate/

1 Introduction

We are investigating the relation between the word order of languages and the speakers' thought pattern. We have approached it through suicide rate and homicide rate (Ehara, 2015, 2017). Previous works concluded that head final language speakers tend to have higher suicide rate and lower homicide rate and head initial language speakers tend to have lower suicide rate and higher homicide rate. Higher suicide rate relates to intropunitive thought pattern and higher homicide rate relates to extrapunitive thought pattern.

In this paper we use another metric concerning to a thought pattern. It is the entropy of Y-chromosome haplogroup (YHg) distribution of the speakers' population (shortly "entropy"). Higher entropy value means the population is rich in diversity and it relates to peaceful thought pattern. Lower entropy value mean the population is poor diversity and it relates to warlike thought pattern (Sakitani, 2008).

Our conjecture is that head final language speakers' population has higher entropy value and head initial language speakers' population has lower entropy value.

2 Data

Data for the word order features are obtained from the WALS online database (Dryer, 2013). We use two dominant word order features:

- Order of Object (O) and Verb (V),
- Order of Adjective (A) and Noun (N).

WALS online provides O and V feature values for 1519 languages and A and N feature values for 1366 languages.

YHg data of populations are obtained from Wikipedia's page of "Y-chromosome haplogroups by populations" (Wikimedia Foundation, 2015). After preprocessing (Ehara, 2016), we get the YHg data from 452 populations. The number of languages are 196 in this data. In the Wikipedia pages, the granularity of YHg is different page by page. We adjust the granularity to the coarsest 20 groups from A to T (Karafet et al., 2008). The entropy is calculated by

$$H = \sum_{g \in G} p(g) \log_{10} p(g)$$

where G is the set of YHg $= \{A, B, \ldots, T\}$ and $p(g)$ is the probability (relative frequency) of YHg g in a population. Appendix 1 shows the base data used in the research sorted by the entropy.

3 Analysis and results

We conduct t-test for O and V word order groups and A and N word order groups. We discard "no dominant order" data. Results are shown in Table 1. Both O and V case and A and N case, mean values are different significantly. The entropy of OV type language speakers' population is tend to be higher than that of VO type language speakers' population. The entropy of AN type language speakers' population is tend

to be higher than that of NA type language speakers' population. This results shows our conjecture is significantly true.

Table 1: Results of t-test for the entropy

	OV	VO
n	83	82
mean	0.453	0.363
standard dev.	0.209	0.192
t	2.909	
p	0.0041	

	AN	NA
n	79	86
mean	0.486	0.348
standard dev.	0.191	0.199
t	4.533	
p	0.00001	

4 Distance from Africa

Homo sapiens was born in Africa and spread to all over the world. Because the diversity of haplogroup is due to the change in base sequence, the diversity is rich in near location from Africa and poor in far location from Africa. We consider it in our analysis using the linear regression. We measure the geometrical distance of languages from Africa by the great circular distance (GCD) calculated by the position data (latitude and longitude) in the WALS. The origin is set to the Amharic position. GCD values are listed in Appendix 1. From the result of linear regression analysis, we obtain the regression formula calculating the entropy (H) from GCD :

$H = -0.0711 * GCD + 0.4819$

Scattering graph between GCD and H of languages is shown in Figure 1 with the liner regression line.

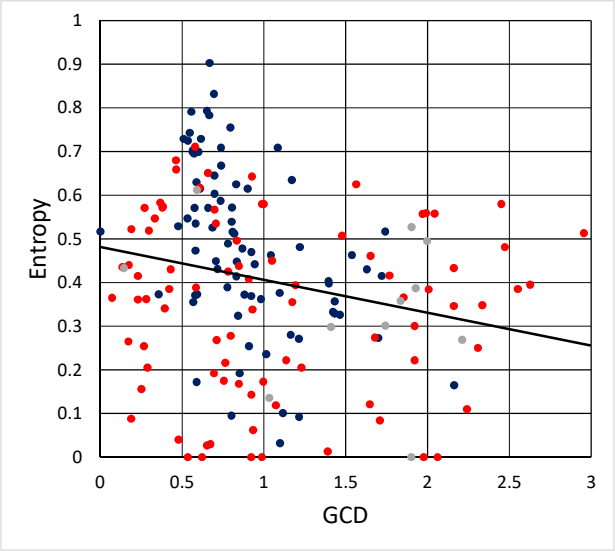

Figure 1: Linear regression result (red:NA, blue:AN, gray:No dominant order or No data)

We conduct again t-test for the residuals of the regression. The results are shown in Table 2. We recognize again OV and VO groups and AN and NA groups both have the significantly different mean values for the residuals.

Table 2: Results of t-test for the residuals

	OV	VO
n	83	82
mean	0.040	−0.044
standard dev.	0.196	0.198
t	2.750	
p	0.0066	

	AN	NA
n	79	86
mean	0.067	−0.058
standard dev.	0.181	0.205
t	4.159	
p	0.00005	

5 Conclusion

Relation between word order (Object and Verb and Adjective and Noun) of languages and the entropy of Y-chromosome haplogroup distribution of the speakers' population is examined. T-test results show OV and AN word order language speakers' population tend to have higher entropy value than VO and NA word order language speakers' population.

One of the remaining issues is to clarify the relation between word order and the entropy of the speakers' mitochondrial DNA haplogroup distribution which reflects diversity of populations with female line.

References

Matthew S. Dryer. 2013. Order of Object and Verb and Order of Adjective and Noun, *In: Dryer, Matthew S. & Haspelmath, Martin (eds.) The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/83 and 87, Accessed on 2015-3-23).

Terumasa Ehara. 2015. Relation between Word Order Parameters and Suicide / Homicide Rates, *Journal of Yamanashi Eiwa College*, Vol.13, pages 9-29.

Terumasa Ehara. 2016. Structural Distance, Lexical Distance and Speaker's Genetic Distance of Languages, *Proceedings of The 22th Annual Meeting of The Association for Natural Language Processing*, P5-1, pages 123-126.

Terumasa Ehara. 2017. Relation between the Word Order Characteristics and Suicide/Homicide Rates (6), *Proceedings of The 23th Annual Meeting of The Association for Natural Language Processing*, P4-4, pp.190-193.

Tatiana M. Karafet, Fernando L. Mendez, Monica B. Meilerman, Peter A. Underhill, Stephen L. Zegura, and Michael F. Hammer. 2008. New Binary Polymorphisms Reshape and Increase Resolution of the Hu-

man Y Chromosomal Haplogroup Tree, *Genome Research*, Vol. 18, pages 830-838.
Mitsuru Sakitani. 2008. A long journey of Japanese populations revealed by DNA analysis, *Showado, Kyoto*, pages 37-39 (in Japanese).
崎谷満. 2008. DNAでたどる日本人10万年の旅, *昭和堂, 京都*, pages 37-39.
Wikimedia Foundation. 2015. Y-chromosome haplogroups by populations, *Wikipedia*. (https://en.wikipedia.org/wiki/Y-chromosome_haplogroups_by_populations, Accessed on 2015-11-20).

Appendix 1 Base data used in the research

Feature values for O and V are 1:OV, 2:VO, 3:No dominant order, blank: No data.
Feature values for A and N are 1:AN, 2:NA, 3:No dominant order, blank: No data.
GCD means great circular distance of languages originated by Amharic position.
n means Number of sample.
A to T means Y-chromosome haplogroup name (cell values are relative frequency of YHg in the sample).

No.	WALS language name	O and V	A and N	GCD	n	Entropy	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
36	Emberá (Northern)	1	2	1.97703	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	81	0	0	0
43	Fongbe	2	2	0.62183	100	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
62	Igbo	2	2	0.53436	209	0	0	0	0	0	89.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
65	Irish	2	2	0.98797	222	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82	0	0
130	Rama	1	2	2.06235	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
154	Ticuna	3		1.90058	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
172	Wolof	2	2	0.9221	173	0	0	0	0	0	95.56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	Paiwan	2	2	1.38985	492	0.013	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0	87.9	0	0	0	0	0
40	Ewe	2	2	0.65208	91	0.027	0	0	0	0	95.6	0	0	0	0	0	0	0	0	0	0	0	0	1.1	0	0
110	Mooré	2	2	0.67269	155	0.03	0	1.25	0	0	96.79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
118	Nenets	1	1	1.09876	148	0.032	0	0	0	0	0	0	0	0	0	0	0	0	0	97.3	0	0	1.4	0	0	0
87	Kongo	2	2	0.47796	54	0.04	1.87	0	0	0	98.13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
171	Welsh	2	2	0.93425	88	0.062	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	90	0	0
53	Guaraní	2	2	1.7091	113	0.084	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92.69	4.7	0	0
79	Kikuyu	2	2	0.18887	42	0.088	2	2	0	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
120	Nganasan	1	1	1.21526	38	0.092	0	0	5.3	0	0	0	0	0	0	0	0	0	0	92.1	0	0	0	0	0	0
19	Bashkir	1	1	0.80153	43	0.095	0	0	0	0	0	0	2.3	0	0	0	0	0	0	2.3	0	0	0	95.3	0	0
73	Ket	1	1	1.11597	48	0.101	0	0	6.2	0	0	0	0	0	0	0	0	0	0	0	0	0	93.7	0	0	0
114	Mixtec (Chalcatongo)	2	2	2.24156	28	0.11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	7	0	0
152	Thai	2	2	1.07283	34	0.119	0	0	2.9	2.9	0	0	0	0	0	0	0	0	0	0	88.2	0	0	0	0	0
85	Kaingang	1	2	1.64699	51	0.121	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	8	0	0
2	Achang	1	3	1.03224	40	0.136	0	0	5	0	0	0	0	0	0	0	2.5	0	0	0	92.5	0	0	0	0	0
106	Mandinka (Gambian)	1	2	0.9231	39	0.143	5	3	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80	Kinyarwanda	2	2	0.25138	163	0.156	0	10.34	0	0	88.81	0	0	0	0	0	0	0	0	0	0	0	0	0.42	0	0
67	Jakaltek	2	1	2.16322	71	0.165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87.3	12.7	0	0
20	Basque	1	2	0.84731	176	0.168	0	0	0	0	1.39	0	0	0	2.86	3.43	0	0	0	0	0	0	0	90.95	0	0.56
64	Ingush		1	0.58881	143	0.172	0	0	0	0	0	0	1.4	0	0.7	90.2	0	2.8	0	0	0	0	0	3.5	0	0
48	Gaelic (Scots)	2	2	0.99532	821	0.173	0	0	0	0	0	0	0	0	12.79	0	0	0	0	0	0	0	0	81.41	0	0
173	Xhosa	2	2	0.75569	80	0.175	5	5	0	0	87	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
159	Tamashek	2	2	0.69461	94	0.192	0.54	0.54	0	0	87.2	0	0	0	0	0	1.07	0	0	0	0	0	0	9.58	0	0
166	Udmurt	1	1	0.85166	87	0.192	0	0	0	0	0	0	0	0	1.1	0	0	0	0	85.1	0	0	0	12.6	0	0
105	Mien	2	2	1.22989	70	0.205	0	0	10	1.45	0	0	0	0	0	0	0	0	0	1.45	87.2	0	0	0	0	0
146	Swahili	2	2	0.2885	1108	0.205	0.99	6.25	0	0	87.54	0	0	0	0	0	0	0	0	0	0	0	0	4.03	0	0.18
96	Ladakhi	1	2	0.76333	87	0.216	0	0	1.1	0	0	0	0	2.3	0	0	0	0	0	0	86.2	0	0	10.3	0	0
58	Hmong Njua	2	2	1.13594	169	0.222	0	0	8.9	3.6	0	0	0	0	0	0	0	0	0	1.2	87.6	0	0	0	0	0
86	Koasati	1	2	1.92095	36	0.222	0	0	2.8	0	0	0	0	0	0	0	0	0	0	0	0	0	75	11.1	0	0
93	Khanty	1	1	1.0143	47	0.236	0	0	0	0	0	0	0	0	0	0	0	0	0	76.6	0	0	0	23.4	0	0
102	Maori	2	2	2.30804	124	0.25	0	0	65.58	0	0	0	0	0	0	0	5.06	0	0	0	5.04	0	0	1.58	0	0
37	English	2	1	0.9089	215	0.254	0	0	0	0	0	0	0	0	25	0	0	0	0	0	0	0	0	66.9	0	0
138	Sandawe	1	2	0.26694	68	0.254	4	14	0	0	77	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
144	Somali	1	2	0.17217	201	0.265	0.5	1	0	0	83.1	0	0	0	0	3	0	0	0	0	0	0	0	1	0	10.4
176	Zulu	2	2	0.71095	509	0.268	4.91	15.06	0	0	78.44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Apache (Western)	1		2.21172	96	0.269	0	0	14.6	0	0	0	0	0	0	0	0	0	0	0	0	0	78.1	5.2	0	0
109	Mandarin	2	1	1.21387	698	0.271	0	0	7.6	0.73	0	0	0	0	0	0.58	0.86	0	0	3	83.68	0.14	0.72	0.58	0	0
103	Maung	2	1	1.69936	60	0.273	0	0	63.3	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
126	Pitjantjatjara	1	2	1.67801	35	0.274	0	0	68.7	0	0	0	0	0	0	0	17.1	0	0	0	3	0	0	0	0	0
17	Berber (Middle Atlas)	2	2	0.79733	943	0.278	1.69	1.49	0	0	81.09	0.43	1.06	0	0	4.66	0.1	0	0	0	0	0	0	4.28	0	0
61	Icelandic	2	1	1.16353	181	0.28	0	0	0	0	0	0	0	0	34.2	0	0	0	0	0	0	0	0	65.2	0	0
148	Tagalog	2	3	1.40922	165	0.298	0	0	3.48	0	0	0	0	0	0	0	16.97	0	0	0	74.24	0	0	0	1.39	0
141	Seminole	1	2	1.92101	20	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	50	0	0
122	Oneida	2		1.74232	30	0.301	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	46.7	0	0
98	Latvian	2	1	0.84187	114	0.324	0	0	0	0	0.9	0	0	0	0	0	0	0	0	42.1	0	0	0	39.5	0	0
88	Korean	1	1	1.46489	1001	0.326	0	0.2	11.98	2.17	0	0	0	0	0	0.1	0.4	0.3	0	3.01	78.04	0.1	0.89	0.2	0	0
174	Yakut	1	1	1.43318	222	0.329	0	0	13.49	0	0	0	0	0	1.36	0	0	0	0	72.01	0	0	0	8.06	0	0
39	Even	1	1	1.42212	31	0.333	0	0	74.2	0	0	0	0	0	3.2	0	0	0	0	12.9	0	0	0	6.5	0	0
78	Khasi	2	2	0.93076	92	0.338	0	0	0	0	0	10.9	0	6.5	0	0	0	0	0	0	72.8	4.4	0	0	0	0
72	Kera	2	2	0.39393	54	0.341	1.8	3.7	0	0	24.1	0	0	0	0	0	0	0	0	0	0	0	0	70.4	0	0
116	Navajo	1	2	2.16113	385	0.346	0	0	13.49	0	0	0	0	0	0	0	0	0	0	0	0	0	69.91	14.54	0	0
89	Kiribati	2	2	2.33524	32	0.348	0	0	18.7	0	0	0	0	0	0	0	65.6	0	0	0	9.4	0	0	0	0	0
133	Rutul	1	1	0.56852	24	0.355	0	0	0	0	0	58	37.5	0	0	4.2	0	0	0	0	0	0	0	0	0	0
170	Vietnamese	2	2	1.17392	70	0.355	0	0	4.3	2.9	0	0	0	0	0	2.9	0	0	0	2.9	78.6	0	7.1	0	0	0
108	Manchu	1	1	1.43273	136	0.357	0	0	19.09	0.75	0	0	0	0	0	0	2.23	0	0	3.68	73.61	0.74	0	0	0	0
26	Chipewyan	1		1.83537	48	0.358	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	31	62.5	0	0
101	Maasai	2	2	0.22954	26	0.361	27	8	0	0	66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
95	Komi-Zyrian	2	1	0.98196	197	0.362	0	0	0	0	0	0	0	0	4.07	0	0	0	0	50.75	0	0	0	40.13	0	0
112	Masalit	1	2	0.27952	32	0.362	18.8	3.1	0	0	71.9	0	0	0	0	6.3	0	0	0	0	0	0	0	0	0	0
123	Oromo (Harar)	1	2	0.07103	78	0.365	10.3	1.3	0	0	64.1	0	0	0	0	3.8	0	0	0	0	0	0	0	0	0	5.1
8	Apurinã	3	2	1.85346	44	0.366	0	0	4.55	0	0	0	0	0	0	0	0	0	0	0	0	0	50.25	34.07	0	0
42	Finnish	2	1	0.92252	38	0.369	0	0	0	0	0	0	0	0	28.9	0	0	0	0	63.2	0	0	0	7.9	0	0
97	Lak	1	1	0.57869	314	0.371	0	0	0	0	0.97	2.24	1.6	0	6.07	77.12	0	0	0	0	0.96	0	0	10.97	0	0
45	Frisian	3	1	0.88183	188	0.372	0	0	0	0	2.05	0	0	0	31.5	3.7	0	0	0	0	0	0	0	62.85	0	0
24	Chechen	1	1	0.59242	699	0.373	0	0	0	0	0	0.87	5.41	0	0.28	75.26	0	7.02	0	0.29	0	0	0	9.11	0	0.14
136	Sango	2	1	0.35677	115	0.373	3.47	31.3	0	0	58.24	0	0	0	0	0	0	0	0	0	0	0	0	1.72	0	0
139	Selkup	1	1	1.09644	131	0.376	0	0	1.5	0	0	0	0	0	0	0	0	0	0	6.9	0	0	66.4	25.2	0	0

No.	WALS language name	O and V	A and N	GCD	n	Entropy	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
100	Lakhota	1	2	2.00711	106	0.384	0	0	5.5	0	0	0	0	0	0	0	0	0	0	0	0	0	39.62	42.45	0	0
46	Fula (Cameroonian)	2	2	0.42179	63	0.385	3.17	0	0	0	60.27	0	0	0	0	0	0	0	0	0	0	0	0	22.2	0	4.76
158	Tongan	2	2	2.55251	55	0.385	0	0	23	0	0	0	0	0	0	0	1	0	8	0	60	0	0	0	0	0
29	Cree (Plains)	2		1.92881	359	0.387	0	0	6.39	0	0	0	0	0	0	0	0	0	0	0	0	0	29.17	45.97	0	0
69	Juǀ'hoan	2	2	0.5846	64	0.388	36	8	0	0	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
127	Polish	2	1	0.77636	1006	0.389	0	0	0	0	4.08	0	0	0	17.19	2.27	0	0	0	3.65	0	0	0	68.66	0	0
63	Indonesian	2	2	1.19295	1536	0.394	0	0	15.48	0	0	0.33	0	1.22	0	0	3.66	0	1.21	0	66.44	0	0.14	0	2.61	0
135	Samoan	2	2	2.628	87	0.395	0	0	60.93	0	0	0	0	0	0	0	3.43	0	2.28	0	28.73	0	0	0	1.14	0
38	Evenki	1	1	1.39714	96	0.398	0	0	67.7	0	0	0	0	0	5.2	0	0	0	0	19.8	0	0	4.2	1	0	0
31	Dagur	1	1	1.39488	39	0.403	0	0	30.8	0	0	0	0	0	0	0	0	0	0	10.3	53.8	0	0	0	0	0
49	Garo	1	2	0.90768	71	0.408	0	0	8.5	0	0	4.2	0	1.4	0	0	4.2	0	0	0	74.7	7	0	0	0	0
107	Mari (Meadow)	1	1	0.83139	159	0.414	0	0	0	0	0	0	0	0	5.65	1.9	0	0	0	44	0	0	0	47.14	0	0
47	Fur	1	2	0.23033	32	0.415	31.3	3.1	0	0	59.4	0	0	0	0	6.3	0	0	0	0	0	0	0	0	0	0
92	Koryak	1	1	1.71994	27	0.415	0	0	59.2	0	0	0	0	0	0	0	0	0	0	22.2	0	18.5	0	0	0	0
6	Arrernte (Mparntwe)	1	2	1.76791	108	0.416	0	0	66.2	0	0	3	0	0	0	0	22.2	0	0	0	0.9	0	0	8	0	0
30	Catalan	2	2	0.78125	250	0.425	0	0	0	0	9.55	0	5.41	0	6.14	0.4	0	0	0	0	0	0	0	67.16	0	4.3
34	Doyayo	2	2	0.42965	72	0.43	1.4	12.5	0	0	58.4	0	0	0	0	0	0	0	0	0	0	0	0	27.8	0	0
68	Japanese	1	1	1.62985	493	0.43	0	0	6.88	36.9	0	0	0	0	0	0	0	0	0	2.05	53.19	0	0.4	0	0	0
140	Slovene	2	1	0.7157	75	0.431	0	0	0	0	2.7	0	2.7	0	30.7	4	0	0	0	0	0	0	0	60	0	0
14	Beja	1	3	0.14367	42	0.433	4.8	0	0	0	52.4	0	0	0	0	38.1	0	0	0	0	0	0	0	4.8	0	0
94	Kwaio	2	2	2.16138	32	0.433	0	0	0	0	0	0	0	0	0	0	59.4	0	9.4	0	28.1	0	0	0	3.1	0
111	Moro	2	2	0.13555	28	0.435	46.4	14.3	0	0	39.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	French	2	2	0.8471	23	0.438	0	0	0	0	8.7	0	0	0	17.4	4.3	0	0	0	0	0	0	0	52.2	0	0
33	Dinka	3	2	0.17423	107	0.44	53.37	24.29	0	0	22.34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
134	Romani (Welsh)	2	1	0.94294	57	0.442	0	0	0	0	29.8	0	0	59.6	5.3	1.8	0	0	0	0	0	0	0	3.6	0	0
51	German	3	1	0.83365	1231	0.448	0	0	0	0	6.2	0	0	0	23.58	3.95	0	0	0	1.58	0	0	0	56.92	0	0
77	Khoekhoe	1	1	0.7067	413	0.449	42.13	12.11	0	0	43.1	0	0	0	0	0.24	0	0	0	0	0	0	0	0.71	0	0
15	Batak (Karo)	2	2	1.04981	57	0.45	0	0	5.3	1.8	0	14	0	0	0	0	3.5	0	0	0	61.3	0	0	0	3	0
104	Maybrat	2	2	1.65217	245	0.461	0	0	16.29	0	0	0	0	0	0	0	6.97	0	64.92	0	1.99	0	0	0	9.38	0
143	Saami (Northern)	2	1	1.04193	38	0.463	0	0	0	0	0	0	0	0	31.6	0	0	0	0	44.7	0	0	0	23.7	0	0
165	Udihe	1	1	1.53759	76	0.463	0	0	42.12	0	0	0	0	0	0	0	1.31	0	0	14.49	42.13	0	0	0	0	0
147	Swedish	2	1	0.92273	160	0.47	0	0	0	0	1.3	0	0	0	37.5	0	0	0	0	14.4	0	0	0	37.5	0	0
12	Avar	1	1	0.58191	177	0.473	0	0	0	0	1.68	0	7.39	0	1.17	66.18	0	3.94	0	1.1	0	0	0	13.58	0	1.14
35	Dutch	3	1	0.86823	410	0.478	0	0	0	0	2.9	0	4.1	0	32.9	5.1	0	0	0	0.2	0	0	0	53.5	0	0
23	Buriat	1	1	1.2194	238	0.481	0	0	63.9	0	0	0	0.4	0	0.4	0	8.8	0	0	20.2	0	1.7	1.7	2.9	0	0
163	Tuvaluan	1	2	2.47249	100	0.481	0	0	17	0	0	2	0	0	0	0	36	0	0	0	45	0	0	0	0	0
16	Belorussian	3	1	0.78052	68	0.489	0	0	0	0	4.4	0	1.5	0	25	0	0	0	0	8.8	0	0	0	50	0	0
156	Tolai	2	3	1.99723	395	0.495	0	0	2.3	0	0	0	0	0	0	0	46.8	0	39	0	7.1	0	0	0	4.6	0
145	Spanish	2	2	0.83355	940	0.496	0	0	0	0	13.64	0	3.46	0	7.76	9.64	0	0	0	0	0	0	0	57.97	0	0
121	Ngad'a	2	2	1.47874	71	0.507	0	0	63.3	0	0	0	0	0	0	0	11.3	0	2.8	1.4	8.5	0	0	0	12.7	0
164	Tatar	1	1	0.81811	76	0.512	0	0	0	0	2.6	0	11.8	0	0	47.4	0	0	0	0	1.3	0	0	36.8	0	1.3
149	Tahitian	2	2	2.95615	810	0.513	0	0	44.68	0	0	0	0	0	0	0	12.21	0	6.8	0	28.42	0	0	0	0.11	0
113	Mundari	1	1	0.80939	1681	0.516	0	0	0	0	0	3.67	0	24.18	0	4.13	0	0	0	0	56.17	1.66	0	7.5	0	0
5	Amharic	1	1	0	290	0.517	17.27	1.02	0	0	46.92	0	0	0	0	27.96	0	0	0	0	0	0	0	0	0	3.78
25	Chukchi	1	1	1.74162	24	0.517	0	0	4.2	0	0	0	0	0	0	0	0	0	0	58.3	0	20.8	15.5	4.2	0	0
28	Coptic	2	2	0.29634	66	0.519	0	7.6	0	0	21.2	0	0	0	0	45.5	0	0	0	0	0	0	0	15.2	0	0
117	Nubian (Dongolese)	1	2	0.1891	78	0.522	0	3.85	0	0	23.1	0	0	0	2.55	43.6	0	0	0	0	0	0	0	10.3	0	0
167	Ukrainian	2	1	0.68458	103	0.526	0	0	0	0	1.94	0	1.94	0	21.34	7.75	0	0	0	5.85	0	0	0	58.26	0	0.97
175	Yup'ik (Central)	3	3	1.9026	149	0.527	0	0	0.68	0	0	0	0	0	0	0	0	0	0	27.97	0	4.03	40.93	17.45	0	0
76	Khalaj	1	1	0.47634	136	0.529	0	0	4.45	0	0	0.75	0	0	2.21	0	2.95	0	0	36.04	0	4.44	0	49.27	0	0
124	Ossetic	1	1	0.58316	533	0.535	0	0	0	0	0.93	0.94	60.39	0	2.44	14.06	2.25	0.58	0	0.17	0	0.56	1.32	10.31	0	0.19
150	Tajik	1	2	0.70658	38	0.535	0	0	2.6	0	3	0	0	5	0	18.4	0	8	0	0	0	0	0	52.6	0	0
132	Russian	2	1	0.80285	995	0.539	0	0	0	0	3.33	0	0.5	0	14.39	2.5	0	0	0	26.51	0	0	0	48.46	0	0.29
10	Arabic (Gulf)	2	2	0.3332	203	0.547	0	0	0	0	11.8	0	2.5	0	0.5	57.6	0	1	0	1	0	0	0	17.7	0	5.9
55	Hausa	2	1	0.53198	32	0.547	12.5	15.6	0	0	15.6	0	0	0	0	0	0	0	0	0	0	0	0	40.6	0	0
151	Tigak	2	2	1.97081	109	0.557	0	0	8.3	0	0	0	0	0	0	0	29.4	0	48.6	0	8.3	0	0	0	5.5	0
54	Halia	2	2	2.04431	75	0.558	0	0	1.3	0	0	5.3	0	0	0	0	36	0	41.3	0	14.7	0	0	0	1.3	0
82	Kilivila	2	2	1.99055	53	0.559	0	0	9.4	0	0	0	0	0	0	0	22.6	0	30.2	0	37.7	0	0	0	0	0
66	Italian	2	2	0.69855	323	0.567	0	0	0	0	9.83	0	15.24	0	28.96	0	0	0	0	0	0	0	0	26.5	0	0
7	Arabic (Modern Standard)	2	2	0.27007	4564	0.571	0.18	0.11	0	0	41.91	0.64	2.88	0	1.57	36.92	0.42	1.32	0	0	0	0	0	7.49	0	1.87
60	Hunzib	1	1	0.57545	104	0.571	0	0	0	0	0.94	0	7.68	0	14.4	48.07	0	0.96	0	0	0	0	0	24.99	0	0.94
137	Serbian-Croatian	2	1	0.65927	725	0.571	0	0	0	0	15.16	0	1.64	0	45.36	5.26	0	0.15	0	1.78	0	0	0	25.54	0	0
27	Chuvash	1	1	0.80479	79	0.572	0	0	0	0	0	0	0	0	11.3	24.2	0	0	0	27.8	0	0	0	35.4	0	0
56	Hebrew (Modern)	2	2	0.37873	156	0.572	0	0	0	0	19.2	0	0	0	11.5	28.2	0	0	0	0	0	0	0	33.4	0	0
115	Neo-Aramaic (Arbel Jewish)	1	2	0.38269	48	0.574	0	0	0	0	4.2	0	8.3	0	0	29.2	0	0	0	0	0	0	0	37.5	0	8.3
18	Burmese	1	2	0.98911	159	0.58	0	0	8.17	0.63	0	14.15	0	0	0	0	22.01	0	0	2.52	43.39	0	0	0	0	0
32	Digaro	1	2	0.99843	100	0.58	0	0	0	20.8	0	0	0	0	0	0.8	0	0	0	34.6	35.3	0	0.8	7.7	0	0
41	Fijian	2	2	2.45174	1379	0.58	0	0	10.41	0	0	0	0	0	0	0	32.49	0	41.59	0	7.84	0	0	0	5.72	0
3	Arabic (Egyptian)	2	2	0.3672	536	0.583	0.09	0.09	0	0	43.64	1.49	4.34	0	0.83	31.4	0.75	0	0	0	0	0	0	5.91	0	4.24
142	Sinhala	1	1	0.73501	39	0.587	0	0	0	0	0	10.3	0	10.3	0	10.3	0	18	0	0	0	0	0	51.3	0	0
59	Hungarian	2	1	0.69857	260	0.603	0	0	0	0	6.58	0	3.85	4.23	25.38	8.84	0	0.38	0	0.39	0	0	0	50.37	0	0
70	Kabardian	1		0.59062	827	0.612	0	0	0.96	0	0.95	1.69	52.6	0.12	2.64	13.61	1.34	0.95	0	1.32	0	0.49	0.36	14.75	0	0.24
83	Kalmyk	1	1	0.90008	128	0.615	0	0	59.4	1.55	0	0.8	0	0	0	0.8	4.68	0.8	0	3.09	9.33	8.61	0	10.91	0	0
155	Turkmen	1	1	0.60846	30	0.615	0	0	0	0	0	13	0	0	0	17	13	0	0	0	0	10	0	47	0	0
4	Albanian	2	2	0.608	314	0.617	0	0	0	0	35.37	0	1.28	0	14.93	20.39	0	0	0	0	0	0.64	0	25.2	0	0
119	Nepali	1	1	0.83045	77	0.625	0	0	7.8	0	0	0	0	11.7	0	10.4	0	0	0	0	20.8	0	1.3	45.5	0	0
160	Tetun	2	2	1.56401	39	0.625	0	0	43.5	0	0	0	0	0	0	0	17.9	0	7.7	0	17.9	0	0	0	12.8	0
90	Karachay-Balkar	1	1	0.58783	38	0.63	0	0	0	0	2.6	0	28.9	2.6	2.6	7.9	0	5.3	0	0	0	0	0	34.3	0	0
75	Khalkha	1	1	1.17079	260	0.635	0	0	53.25	1.89	0	0.78	0	0	0	1.56	2.66	0	0	9.19	17.63	2.69	1.12	5.72	0	0
153	Tibetan (Standard Spoken)	1	2	0.92718	1382	0.643	0	0	7.6	25.38	0	4.99	0	0.21	0	0.14	14.23	0	0	0.73	43.39	0.65	0.36	0.51	0	0
57	Hindi	1	1	0.69661	205	0.645	0	0	2.4	0	0	2.4	0.5	28.8	0	10.7	0	3.9	0	0	4.9	0	1	45	0	0
131	Romanian	2	2	0.65756	634	0.651	0	0	0	0	15.47	0	3.47	0	28.84	16.87	0	0	0	0.64	0	0	0.17	32.83	0	0.33
91	Kurdish (Central)	1	2	0.46364	120	0.659	0	0	0	0	7.53	0	3.33	0	13.3	35	0	7.53	0	0	0	0	0	29.98	0	0
22	Burushaski	1	1	0.73823	97	0.668	0	0	8.2	0	0	1	1	4.1	0	8.2	0	47.9	0	0	3.1	1	2.1	10.9	0	0
128	Persian	1	2	0.46274	1055	0.68	0	0	0	0	8.58	0	11.82	0	0.44	31.8	0	5.11	0	0.19	0	0	0	24.16	0	3.4
99	Lezgian	1	1	0.57294	192	0.695	0	0	1.57	0	2.62	7.79	11.95	0	1.57	40.61	3.65	0.51	0	0	0	0	0	28.09	0	1.05
21	Bulgarian	2	1	0.60098	1377	0.699	0	0	0.29	0	21.32	0	4.13	0.42	28.46	14.09	0	0.19	0	0.44	0	0	0.23	27.5	0	1.45
52	Greek (Modern)	2	1	0.56428	681	0.699	0	0	0	0	23.8	0	6.92	0	16.9	20.44	0	0.14	0	0	0	0	0	27.02	0	1.94
50	Georgian	1	1	0.56603	385	0.703	0	0	0	0	1.88	9.61	26.99	0	1.55	30.91	0.52	1.58	0	0	0	0.52	0	21.54	0	1.58
71	Kashmiri	2	1	0.73703	59	0.709	0	0	0	0	0	0	13.57	18.59	0	11.87	0	18.64	0	0	1.78	0	0	27.05	0	0
162	Tuvan	1	1	1.08484	113	0.709	0	0	7.1	0	0	3.5	0.9	0	0.9	0	8.9	0	0	23.9	0	35.4	0	17.7	0	0
1	Abkhaz	1	2	0.57922	172	0.711	0	0	0.57	0	2.88	2.36	34.35	0	4.04	10.45	1.14	2.32	0	0	0	0	1.74	26.08	0	0.57
11	Armenian (Eastern)	3	1	0.53474	774	0.725	0	0	0	0	4.93	3.1	9.67	0	3.21	32.75	0.39	1.91	0	0.23	0	0.26	0.11	35.05	0	8.67
129	Pashto	1	1	0.61469	176	0.729	0	0	1.09	0	0	1.15	7.97	3.99	0	9.58	0.53	11.98	0	0	2.84	0.53	7.96	47.65	0	0.55
161	Turkish	1	1	0.5083	148	0.729	0	0	0	0	13.55	0	11.5	0	19.62	16.87	0	0	0	1.36	0	0	0	31.73	0	2.72
168	Urdu	1	1	0.54713	176	0.743	0	0	7.4	0	0	0	6.2	6.2	0	15.3	0	13.1	0	0	2.3	0	3.4	39.2	0	0
74	Kirghiz		1	0.79445	492	0.755	0	0	33.33	0.78	0	1.87	1.21	0	1.02	2.64	0.81	0	0	3.44	8.76	6.31	3.64	33.12	0	0.19
81	Karakalpak	1	1	0.66612	44	0.783	0	0	22.7	0	0	9	0	0	0	9.1	6.8	5	0	2.3	11.4	0	0	34.1	0	0
13	Azerbaijani	1	1	0.55654	72	0.791	0	0	0	0	5.6	11	18.1	0	2.8	30.6	0	6.9	0	0	0	0	0	20.8	0	4.2
84	Kannada	1	1	0.65179	353	0.793	0	0	1.7	0	0	9.3	2.3	32.9	0	10.5	0	11.6	0	0	13.6	0	0.3	17.5	0	0
157	Tamil	1	1	0.69473	91	0.832	0	0	3.3	0	0	9.9	5.5	25.3	0	19.8	0	15.4	0	0	1.1	3.3	0	13.2	0	0
169	Uzbek	1	1	0.66804	503	0.903	0	0	10.35	2.25	1.46	5.75	0.6	0	1.6	11.14	5.94	2.78	0	3.01	6.16	4	0.4	33.16	0	0