Relation between Word Order Characteristics and Suicide / Homicide Rates (6) — Adding a Religious Feature —
語順特徴と自殺率／他殺率との関係(その 6) — 宗教素性の追加 —

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

In previous papers (Ehara, 2010, 2011, 2012, 2013, 2015b), we have investigated quantitative relations between the word order characteristics and suicide / homicide rates. The purpose of our study is to clarify relations between syntactic structures of a language, especially word order structures, and people's thought pattern who use it as a native language.

In the previous papers, we used non-linguistic features: economic feature (GDP per capita and Gini coefficient) and climate features (average annual temperature and average annual precipitation) in addition to linguistic features (order of object and verb and order of adjective and noun) to analyze the relation. In this paper, we add a religious feature.

2 Data

Data for the word order characteristics (features) are obtained from the WALS database (Dryer, 2005). We use two major word order features:

(Fea-1) Order of Object (O) and Verb (V),

(Fea-2) Order of Adjective (A) and Noun (N).
We can get feature values of both two features to 65 languages. OV has +1 and VO has −1 as the feature values for Fea-1. AN has +1 and NA has −1 as the feature values for Fea-2.

Suicide rate and homicide rate are obtained from the WHO's "mortality and burden of disease estimates for WHO member states in 2004" (WHO, 2009). We use the base-10 logarithm values of the suicide rate and the homicide rate as the feature values. They are named as S-rate and H-rate.

We use two economic features. One is base-10 logarithm of GDP per capita with US dollars (GDP) and the other is Gini coefficient (GINI). These economic feature values are gathered mainly from World Bank's data (World Bank, 2014; Ehara, 2015). Population data of countries are also gathered mainly from World Bank (World Bank, 2014; Ehara, 2015).

Average annual temperature (TMP) and average annual precipitation (PRC) data were gathered by the method described in Ehara (2012).

As a religious feature, we use a religious type (REL): Abrahamic religious type and other type. Abrahamic type includes Christian, Muslim and Jewish religions; and other type includes Hindu, Buddhism, indigenous beliefs and other religions. Non-religion is included in the other type. Feature values for the religious feature are determined by the CIA's world fact book (CIA, 2004). Most popular religion for each country is determined as the religion of the country. Feature value for Abrahamic type is −1 and feature value for other type is +1.

We merge S-rate, H-rate, GDP, GINI, TMP, PRC and REL feature values from country based to language based by the method of Ehara (2013).

Appendix 1 shows the base data we used in this paper[1].

3 Analysis and results

We make multiple regression analysis. Criterion variables are S-rate and H-rate. Explanatory variables are GDP, GINI, TMP, PRC, REL,

---

[1] Used data are opened at my web site.

Fea-1 and Fea-2.

Results are shown in Table 1 and Table 2. Yellow-colored p-values are less than 5% and blue-colored p-values are less than 1%[2].

Table 1: Result of multiple regression analysis for S-rate

S-rate	Partial regression coefficient	Standardized partial regression coefficient	T-value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	0.0066	0.0128	0.0859	0.4660	0.2327	0.0114
GINI	-0.0036	-0.0779	-0.6249	0.2673	-0.1730	-0.0825
TMP	-0.0112	-0.2216	-1.2698	0.1047	-0.2678	-0.1659
PRC	0.0018	0.3219	2.2854	0.0130	0.1220	0.2897
REL	0.1149	0.2472	2.0839	0.0209	0.2890	0.2661
Fea-1	-0.0854	-0.2130	-1.5356	0.0651	-0.0835	-0.1993
Fea-2	0.1307	0.3527	2.2675	0.0136	0.3676	0.2876
Intercept	1.1010	0.0000	2.8711	0.0029		

Table 2: Result of multiple regression analysis for H-rate

H-rate	Partial regression coefficient	Standardized partial regression coefficient	T-value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.3838	-0.5832	-4.2283	0.0001	-0.4921	-0.4886
GINI	0.0222	0.3763	3.2537	0.0010	0.4580	0.3958
TMP	-0.0018	-0.0285	-0.1763	0.4304	0.3215	-0.0234
PRC	0.0009	0.1332	1.0200	0.1561	0.2266	0.1339
REL	-0.0327	-0.0557	-0.5060	0.3074	0.0605	-0.0669
Fea-1	-0.1201	-0.2366	-1.8396	0.0355	0.0745	-0.2367
Fea-2	0.1012	0.2159	1.4965	0.0700	-0.1495	0.1944
Intercept	1.0322	0.0000	2.2933	0.0128		

From Table 1 and Table 2, we can recognize the following statements.

(1) PRC, REL and Fea-2 have positive partial regression coefficient to S-rate with 5% significance level.

(2) GDP has negative partial regression coefficient to H-rate with 1% significance level, GINI has positive partial regression coefficient to H-rate with 1% significance level and Fea-1 has negative partial regression coefficient to H-rate with 5% significance level

(3) Linguistically, head-final (OV and AN) language speakers have high S-rate and low H-rate and head-initial (VO and NA) language speakers have high H-rate and low S-rate.

These facts may implicate that: head-final language speakers tend to have introverted

thinking pattern and head-initial language speakers tend to have extroverted thinking pattern.

Adding the religious feature, contribution ratio for S-rate rises from 28% to 33% and contribution ratio for H-rate slightly rises from 42% to 43%.

4 Related works

Nabekura (1997) shows experimental results obtained by the Rosenzweig's picture-frustration study. In his study, bilingual speakers of Japanese and English show a tendency to have an extropunitive thought pattern when they use English and they show a tendency to have an intropunitive or impunitive thought pattern when they use Japanese.

Kashima and Kashima (1998, 2003) conducted quantitative studies for the relation between culture types and "pronoun drop". They used Hofstede's four cultural dimensions especially individualism vs. collectivism dimension (Hofstede, 1991). Their statistical analysis shows no pronoun drop language speakers have higher individualism score than pronoun drop language speakers.

Schouwstra et al. (2011) shows in her non-verbal (gestural) communication experiments, OV order is preferred when the verb is extensional (ex. cover) and VO order is preferred when the verb is intensional (ex. want).

Shibata (2014) conducted the studies for the relation between word order and cognitive structures of speakers. He gave his attention to the order of two-digit complex numerals. He concluded that speakers of languages such that tens place precedes ones place (ex. Chinese "shísān") have zoom-in-type (bottom up) cognitive pattern and speakers of languages such that ones place precedes tens place (ex. English "thirteen") have zoom-out-type (top down) cognitive pattern.

Concerning OV and VO word order, Yamamoto (2003) shows by his historical studies that most of the areas around the world were covered

---

[2] In the previous studies, we use two-sided p-value. However, in this study we use one-sided p-value, because we have interest only in the sign of the partial regression coefficients. Explicitly, if the realization (a) of the partial regression coefficient (A) of a feature is positive,

null hypothesis is $A \leqq 0$ and alternative hypothesis is $A > 0$. And in the $a < 0$ case, null hypothesis is $A \geqq 0$ and alternative hypothesis is $A < 0$.

by OV word order language speakers in ancient times.

Other related works were surveyed in Ehara (2015a).

5 Conclusion

We have examined the relation between economic, climate, religious and linguistic features and suicide rate (S-rate) / homicide rate (H-rate). We make multiple regression analysis. Explanatory variables are GDP per capita (GDP), Gini coefficient (GINI), average annual temperature (TMP), average annual precipitation (PRC), religious type (REL) and two word order features (Fea-1: OV / VO and Fea-2: AN /NA).

For the word order features, we can conclude: head-final (OV and AN) language speakers have high S-rate and low H-rate and head-initial (VO and NA) language speakers have high H-rate and low S-rate.

References

Central Intelligence Agency (CIA). 2004. Religions, *The World Factbook*, Field 2122.html.
https://www.cia.gov/library/publications/download/download-2004/factbook2004.zip
(accessed on March 23, 2015).

Matthew S. Dryer. 2005. Word Order, *The World Atlas of Language Structures,* Chapter F, pp.330-397, Oxford University Press, 2005.
http://wals.info/

Terumasa EHARA. 2010. Relation between the Word Order Characteristics and Suicide/Homicide Rates, *Proceedings of The 16th Annual Meeting of The Association for Natural Language Processing,* E4-2, pages 956-959.

Terumasa EHARA. 2011. Relation between the Word Order Characteristics and Suicide/Homicide Rates (2), *Proceedings of The 17th Annual Meeting of The Association for Natural Language Processing,* F4-6, pages 1037-1040.

Terumasa EHARA. 2012. Relation between the Word Order Characteristics and Suicide/Homicide Rates (3), *Proceedings of The 18th Annual Meeting of The Association for Natural Language Processing*, B1-5, pages 54-57.

Terumasa EHARA. 2013. Relation between the Word Order Characteristics and Suicide/Homicide Rates (4), *Proceedings of The 19th Annual Meeting of The Association for Natural Language Processing*, P6-3, pages 826-829.

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

Terumasa EHARA. 2015b. Relation between the Word Order Characteristics and Suicide/Homicide Rates (5), *Proceedings of The 21th Annual Meeting of The Association for Natural Language Processing*, P1-4, pages 16-19.

Geet Hofstede. 1991. *Cultures and Organizations – Software of the mind –*, McGraw-Hill.

Emiko Kashima and Yoshihisa Kashima. 1998. Culture and Language – The Case of Cultural Dimensions and Personal Pronoun Use –, *Journal of Cross-Cultural Psychology*, Vol.29, No. 3, pages 461–486.

Yoshihisa Kashima and Emiko Kashima. 2003. Individualism, GNP, Climate, and Pronoun Drop – Is Individualism Determined by Affluence and Climate, or Does Language Use Play a Role? –, *Journal of Cross-Cultural Psychology*, Vol. 34, No. 1, pages 125–134.

Takeyoshi Nabekura. 1997. *Introduction to Intercultural Communication*, pages 119–140, Maruzen (in Japanese).
鍋倉健悦 1997. *異文化コミュニケーション入門*, pages 119–140, 丸善ライブラリー.

Marieke Schouwstra, Anouschka van Leeuwen, Nicky Marien, Marianne Smit and Henriëtte de Swart. 2011. Semantic structure in improvised communication, *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, pages 1497–1502.

Katsuyuki Shibata. 2014. Elementary Mathematical Education and World Historical Linguistics, *Information Processing Society of Japan Technical Report,* Vol.2014-NL-218, No.7, pages 1-6 (in Japanese).
柴田勝征 2014. 算数教育と世界歴史言語学、*情報処理学会研究報告*, Vol.2014-NL-218, No.7, pages 1-6.

The World Bank. 2014. *World Development Indicators*, Version: 6 November, 2014.
http://data.worldbank.org/data-catalog/world-development-indicators

World Health Organization (WHO). 2009. *Mortality and burden of disease estimates for WHO member states in 2004.*
http://www.who.int/entity/healthinfo/global_burden_disease/gbddeathdalycountryestimates2004.xls

Hideki Yamamoto. 2003. *Survey and Historical Study of Geographical and Genealogical Distribution of Word Order around the World*, pages 137-141, Keisuisha Co., Ltd (in Japanese).
山本秀樹 2003. *世界諸言語の地理的・系統的語順分布とその変遷*, pages 137-141, 渓水社.

Appendix 1 Base data used in this paper

No.	Language name	Language code	No. of countries	Population(10**3)	S-Rate log10(Suicide rate)	H-Rate log10(Homicide rate)	GDP log10(gdp per capita)	GINI Gini coefficient	Fear-1 VO / OV	Fear-2 NA / AN	TMP Average annual temperature(degree C)	PRC Average annual precipitation(cm)	REL Religious type
1	Albanian	alb	1	3014.579	0.881	0.821	3.394	31.090	-1	-1	15.2	145.9	-1.000
2	Amharic	amh	1	74066.147	0.772	1.287	2.128	29.860	1	1	16.6	114.5	-1.000
3	Arabic (Egyptian)	aeg	1	70591.288	0.200	0.105	3.048	32.260	-1	-1	21.7	3.5	-1.000
4	Arabic (Gulf)	arg	4	168091.237	1.062	0.847	3.363	33.060	-1	-1	25.8	207.5	-1.000
5	Arabic (Iraqi)	arq	1	26673.536	1.185	0.866	3.138	28.600	-1	-1	22.7	15.5	-1.000
6	Arabic (Modern Standard)	ams	11	115864.026	0.685	1.063	3.557	35.765	-1	-1	23.7	40.3	-1.000
7	Arabic (Moroccan)	amr	2	32911.245	0.360	0.380	3.246	40.806	-1	-1	18.3	47.7	-1.000
8	Arabic (Syrian)	asy	1	17676.012	-0.280	0.413	3.152	35.780	-1	-1	16.9	17.6	-1.000
10	Azerbaijani	aze	1	8306.500	0.109	0.451	3.019	16.230	1	1	14.2	36.2	-1.000
12	Bulgarian	bul	1	7781.161	1.137	0.478	3.512	28.730	-1	1	10.3	58.6	-1.000
13	Burmese	brm	1	49875.169	0.996	1.197	2.579	38.100	1	-1	27.4	210.8	1.000
14	Catalan	ctl	1	79.060	0.849	-0.080	4.468	28.100	-1	-1	11.0	107.2	-1.000
15	Czech	cze	1	10197.101	1.230	0.126	4.048	26.820	-1	1	8.4	49.9	1.000
16	Danish	dsh	1	5404.523	1.155	-0.020	4.656	24.600	-1	1	9.1	58.2	-1.000
19	English	eng	42	706106.354	0.938	1.037	4.367	40.749	-1	1	18.3	120.7	-0.943
20	Estonian	est	1	1362.550	1.400	0.949	3.946	34.270	-1	1	6.0	67.9	1.000
21	Finnish	fin	1	5228.172	1.320	0.408	4.558	28.270	-1	1	5.3	67.9	-1.000
22	French	fre	18	249364.700	0.997	1.270	3.934	40.373	-1	-1	22.0	117.9	-0.814
23	Georgian	geo	1	4318.300	0.287	0.565	3.074	39.780	1	1	12.7	53.1	-1.000
25	Greek (Modern)	grk	2	12071.549	0.470	-0.086	4.316	33.491	-1	1	18.9	37.1	-1.000
26	Hebrew (Modern)	heb	1	6809.000	0.787	0.676	4.269	41.200	-1	-1	15.9	63.2	-1.000
27	Hindi	hin	1	1110626.108	1.227	0.739	2.813	33.150	1	1	25.2	76.8	1.000
29	Icelandic	ice	1	292.074	1.076	0.005	4.657	27.810	-1	1	4.7	84.7	-1.000
30	Indonesian	ind	1	221293.797	1.031	0.954	3.065	32.590	-1	-1	28.0	182.1	-1.000
31	Irish	iri	1	4070.262	1.092	-0.138	4.661	33.790	-1	-1	9.8	77.5	-1.000
32	Italian	ita	2	57714.617	0.853	-0.019	4.478	37.264	-1	-1	15.4	70.7	-1.000
33	Japanese	jpn	1	127761.000	1.395	-0.270	4.562	32.110	1	1	16.3	152.9	1.000
34	Khalkha	kha	1	2496.248	1.060	0.511	2.902	34.070	1	1	-0.1	28.1	1.000
35	Khmer	khm	1	13149.386	0.652	1.267	2.608	35.530	-1	-1	27.8	140.7	1.000
36	Kinyarwanda	kin	1	9254.379	0.887	1.424	2.354	52.560	-1	-1	20.5	106.0	-1.000
37	Korean	kor	2	71678.718	1.334	0.892	4.035	31.435	1	1	11.9	130.4	-0.340
38	Lao	lao	1	5699.112	1.312	0.729	2.618	33.670	-1	-1	26.7	168.2	1.000
39	Latvian	lat	1	2263.122	1.419	1.009	3.784	34.960	-1	1	6.0	70.2	-1.000
40	Lithuanian	lit	1	3377.075	1.633	0.960	3.827	35.200	-1	1	7.1	63.1	-1.000
41	Macedonian	mcd	1	2085.728	0.939	0.717	3.422	38.850	-1	1	12.0	57.1	-1.000
42	Mandarin	mnd	2	1300241.700	1.228	0.324	3.197	42.532	-1	1	12.9	53.9	1.000
43	Motu	mtu	1	5948.461	0.989	1.181	2.820	50.880	1	-1	26.8	323.6	-1.000
44	Nauruan	nau	1	10.092	0.491	1.078	3.725	66.860	-1	-1	27.8	205.1	-1.000
45	Nepali	nep	1	24921.910	0.995	1.134	2.465	42.260	1	1	19.0	147.6	1.000
46	Norwegian	nor	1	4591.910	1.064	-0.082	4.753	30.170	-1	1	4.8	85.0	-1.000
47	Pashto	psh	1	24018.682	0.796	0.529	2.343	27.800	1	1	13.1	28.9	-1.000
48	Persian	prs	1	69342.126	0.785	0.404	3.372	39.110	1	-1	18.4	24.1	-1.000
49	Polish	pol	1	38182.222	1.230	0.207	3.821	35.430	-1	1	8.4	53.0	1.000
50	Portuguese	por	7	232924.896	0.780	1.466	3.575	54.655	-1	-1	21.3	131.5	-0.863
51	Romanian	rom	1	21451.748	1.106	0.513	3.548	30.040	-1	-1	10.8	60.0	-1.000
53	Russian	rus	1	143821.212	1.563	1.472	3.614	40.770	-1	1	5.8	70.7	-1.000
54	Samoan	sam	1	178.794	0.563	0.054	3.321	46.290	-1	-1	26.5	271.1	-1.000
55	Serbian-Croatian	scr	1	614.670	1.278	0.431	3.528	30.140	-1	1	20.5	1.3	-1.000
56	Sesotho	ses	1	1912.022	0.717	1.122	2.810	51.920	-1	-1	15.5	65.0	-1.000
57	Sinhala	snh	1	19435.000	1.450	0.838	3.027	40.740	1	1	27.7	232.2	1.000
58	Slovene	slo	1	1997.012	1.451	0.309	4.229	31.150	-1	1	10.7	140.1	-1.000
59	Somali	som	1	8249.965	1.398	0.522	2.162	39.700	1	-1	22.5	50.0	-1.000
60	Spanish	spa	20	393990.340	0.807	1.296	3.822	47.996	-1	-1	17.8	93.1	-1.000
61	Swahili	swa	1	37765.139	0.882	1.417	2.544	36.310	-1	-1	26.0	107.2	-1.000
62	Swedish	swe	1	8993.531	1.139	0.092	4.605	26.360	-1	1	6.7	53.6	-1.000
64	Tajik	taj	1	6663.929	0.333	0.343	2.494	33.600	1	-1	15.5	56.8	-1.000
65	Thai	tha	1	65087.400	1.010	0.848	3.394	42.160	-1	-1	28.9	165.3	1.000
66	Tigrinya	tig	1	4665.522	0.694	1.202	2.376	40.000	1	1	16.0	35.4	-1.000
67	Tongan	tng	1	100.319	0.516	-0.197	3.380	37.710	-1	-1	23.5	163.8	-1.000
68	Turkish	tur	1	66845.635	0.562	0.464	3.768	41.290	1	1	12.1	40.8	-1.000
69	Turkmen	tkm	1	4696.876	1.034	0.946	3.163	40.770	1	1	16.8	20.7	-1.000
70	Ukrainian	ukr	1	47451.600	1.430	1.078	3.136	28.930	-1	1	8.4	61.1	-1.000
71	Urdu	urd	1	155151.394	1.013	0.551	2.800	30.920	1	1	23.0	47.0	-1.000
72	Uzbek	uzb	1	25864.350	0.775	0.546	2.668	35.190	1	1	14.8	45.1	-1.000
73	Vietnamese	vie	1	81437.700	0.999	0.577	2.783	36.810	-1	-1	24.5	164.5	1.000