

Relation between Word Order Characteristics and Suicide / Homicide Rates (5)

語順特徴と自殺率／他殺率との関係(その 5)

Terumasa EHARA
江原暉将

Yamanashi Eiwa College
山梨英和大学

<http://www.yamanashi-eiwa.ac.jp/~eharate/>

1 Introduction

In previous papers (Ehara, 2010, 2011, 2012, 2013), 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 climate features (average annual temperature and average annual precipitation) in addition to linguistic features to analyze the relation. In this paper, we add more one economic feature: Gini coefficient. Another new thing is to use economic feature values matched to the measuring year of suicide and homicide rate data.

2 Data

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

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

We can get feature values of both two features to 65 languages.

We define feature value "1" if the order is same as Japanese and "-1" if the order is opposite of "1". For the feature-1 (Fea-1), OV is parameterized to 1 and VO is parameterized to -1. For the feature-2 (Fea-2), AN is parameterized to 1 and NA is parameterized to -1.

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. They are named as S-rate and H-rate.

We use two economic features. One is base-10 logarithm of GDP per capita (GDP) and the

other is Gini coefficient (GINI). These economic feature values are gathered mainly from World Bank's data (World Bank, 2014). In the previous papers, we used economic feature values from the U.N. data. The reason why we change the data source is that the World Bank's data are time series. The measuring year of the S-rate and H-rate is 2004. So, 2004 is desirable for the measuring year of GDP and GINI data. From World Bank's time series data, we can get GDP and GINI value at 2004 point, or at least, we can get interpolated or extrapolated values at 2004 point. Population data of countries are also gathered mainly from World Bank. However, World Bank's data are not exhaustive. We use several other data sources. Table 1 shows the country codes of which we gather economic feature values or population data from other than the World Bank.

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

Table 1: Country codes from other than the World Bank's data

Category	Data source	Country code
GDP	UN, 2011	COK, MMR, NRU, PRK
	PIFS, 2012	NIU
GINI	UN, 2011	AFG, QAT
	Avakov, 2011	KNA, MCO, MHL, SMR
	CIA, 2004	CYP, NZL, PRK, PRT, SGP, VUT, ZWE
	Wikipedia, 2014	ARE, BHR, GNQ, KWT, LBY, OMN, SAU
	PIFS, 2012	COK, KIR, NIU, NRU, PLW, SLB, TON, TUV, WSM
	UNU, 2014	ADO, BHS, BRB, CUB, KOR, LBN, MMR, SOM
	Goodwin, 2009	ATG
	UNDP, 2009	DMA, GRD
	EC, 2008	ERI
	Population	WHO, 2009

We merge GDP, GINI, TMP and PRC data from country based data to language based data by the method of Ehara (2013).

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

3 Analysis and results

We make multiple regression analysis. Criterion variables are S-rate and H-rate. Explanatory variables are GDP, GINI, TMP, PRC, Fea-1 and Fea-2.

At first, we compare the difference of data sources. Table 2 and Table 3 shows comparison results using data from previous data sources (mainly UN data) and new data sources (mainly World Bank data). In these experiments, we do not use GINI data. In the following tables, yellow-colored value of probability means confidence with 1% significance level and purple-colored value of probability means confidence with 10% significance level.

The differences by the data sources are small for both S-rate and H-rate.

Table 2: Comparison result of multiple regression analysis for S-rate

(a) Previous data sources						
S-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.0447	-0.0796	-0.5145	0.6088	0.2124	-0.0668
TMP	-0.0115	-0.2287	-1.2721	0.2083	-0.2675	-0.1634
PRC	0.0022	0.3854	2.8655	0.0058	0.1217	0.3495
Fea-1	-0.0951	-0.2372	-1.6718	0.0999	-0.0839	-0.2127
Fea-2	0.1681	0.4538	2.9514	0.0045	0.3677	0.3587
Intercept	1.0550	0.0000	2.7712	0.0075		

(b) New data sources						
S-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.0082	-0.0157	-0.1043	0.9173	0.2327	-0.0136
TMP	-0.0100	-0.1975	-1.1122	0.2706	-0.2678	-0.1433
PRC	0.0021	0.3744	2.7674	0.0075	0.1220	0.3390
Fea-1	-0.0833	-0.2077	-1.4613	0.1492	-0.0835	-0.1869
Fea-2	0.1642	0.4431	2.8864	0.0054	0.3676	0.3518
Intercept	0.9047	0.0000	2.7245	0.0085		

Table 3: Comparison result of multiple regression analysis for H-rate

(a) Previous data sources						
H-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.4495	-0.6307	-4.1873	0.0001	-0.4993	-0.4786
TMP	-0.0013	-0.0200	-0.1143	0.9094	0.3230	-0.0149
PRC	0.0014	0.2033	1.5523	0.1259	0.2261	0.1981
Fea-1	-0.1269	-0.2496	-1.8072	0.0758	0.0735	-0.2290
Fea-2	0.0643	0.1369	0.9145	0.3642	-0.1502	0.1182
Intercept	2.1170	0.0000	4.5022	0.0000		

(b) New data sources

H-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.4029	-0.6123	-4.1751	0.0001	-0.4921	-0.4776
TMP	0.0001	0.0020	0.0118	0.9907	0.3214	0.0015
PRC	0.0015	0.2125	1.6143	0.1118	0.2266	0.2057
Fea-1	-0.1256	-0.2475	-1.7898	0.0786	0.0745	-0.2269
Fea-2	0.0559	0.1192	0.7983	0.4279	-0.1495	0.1034
Intercept	1.8370	0.0000	4.4928	0.0000		

Next, results with adding GINI data are shown in Table 4 and Table 5. For S-rate, adding GINI data makes no big change. For H-rate, it makes several improvements. GINI gives large contribution to H-rate which is confidence with 1% significance level. Probability that partial regression coefficient of Fea-1 to H-rate be positive reduces from 7.6% to 6.9%.

Table 4: Result of multiple regression analysis for S-rate with adding GINI data

S-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.0135	-0.0260	-0.1713	0.8646	0.2327	-0.0225
GINI	-0.0048	-0.1031	-0.8079	0.4224	-0.1730	-0.1055
TMP	-0.0094	-0.1873	-1.0489	0.2986	-0.2678	-0.1364
PRC	0.0022	0.4006	2.8715	0.0057	0.1220	0.3528
Fea-1	-0.0844	-0.2105	-1.4762	0.1453	-0.0835	-0.1903
Fea-2	0.1560	0.4211	2.6933	0.0092	0.3676	0.3334
Intercept	1.0750	0.0000	2.7275	0.0084		

Table 5: Result of multiple regression analysis for H-rate with adding GINI data

H-rate	Partial regression coefficient	Standardized partial regression coefficient	T-test value	Probability	Correlation coefficient	Partial correlation coefficient
GDP	-0.3779	-0.5743	-4.2247	0.0001	-0.4921	-0.4851
GINI	0.0225	0.3820	3.3399	0.0015	0.4580	0.4016
TMP	-0.0023	-0.0358	-0.2238	0.8237	0.3214	-0.0294
PRC	0.0008	0.1154	0.9228	0.3599	0.2266	0.1203
Fea-1	-0.1204	-0.2372	-1.8563	0.0685	0.0745	-0.2368
Fea-2	0.0941	0.2007	1.4323	0.1574	-0.1495	0.1848
Intercept	1.0389	0.0000	2.3245	0.0236		

Contribution ratios in these experiments are listed in Table 6. Adding GINI, contribution ratio for S-rate rises from 0.2752 to 0.2832 and contribution ratio for H-rate rises from 0.3139 to 0.4246.

4 Conclusion

We examine the relation between economic, climate and linguistic features and suicide rate (S-rate) / homicide rate (H-rate)¹. We make multiple regression analysis. Explanatory variables are GDP per capita (GDP), Gini coeffi-

¹ Used data are opened at my web site.

cient (GINI), average annual temperature (TMP), average annual precipitation (PRC) and two word order features. These word order features are order of Verb and Object noun (Fea-1) and order of Adjective and Noun (Fea-2).

New two things in this paper are (1) adding Gini coefficient as an economic feature and (2) changing data sources for economic feature values and population data.

From the results, we can get similar conclusion of previous papers. However, adding GINI data, probability that partial regression coefficient of Fea-1 to H-rate be positive reduces from 7.9% to 6.9%. We, also, conclude that GINI contributes with positive partial regression coefficient to H-rate with significance level with 1%. Adding GINI data, contribution ratio of explanatory variables to H-rate rises from 0.3139 to 0.4246.

Table 6: Contribution ratio

(a) S-rate

Experimental setting	Contribution ratio
Previous data sources	0.2779
New data sources	0.2752
New data sources + GINI	0.2832

(b) H-rate

Experimental setting	Contribution ratio
Previous data sources	0.3154
New data sources	0.3139
New data sources + GINI	0.4246

References

Alexander V. Avakov. 2011. *Quality of Life, Balance of Power, and Nuclear Weapons*, pp.57-62, Algora Publishing.

Central Intelligence Agency (CIA). 2004. Distribution of family income - Gini index, *The World Factbook*, Field 2172.html. <https://www.cia.gov/library/publications/download/download-2004/>

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 As-*

sociation for Natural Language Processing, E4-2, pp.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, pp.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, pp.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, pp.826-829.

European Commission (EC). 2008. *2007 Annual Operational Review ERITREA*, page 6. http://ec.europa.eu/development/icenter/repository/jar06_er_en.pdf

J. Yolanda Goodwin. 2009. *Millennium Development Goals Report 2009*, page 25. http://www.undp.org/content/dam/barbados/docs/mdg/Antigua_MDG__2010.pdf

Pacific Islands Forum Secretariat (PIFS). 2012. *2012 Pacific Regional MDGs tracking Report*, Figure 2, page 10. <http://www.forumsec.org/resources/uploads/attachments/documents/MDG%20Track%20Rpt%20web%2012.pdf>

United Nations Development Programme (UNDP). 2009. *About Barbados and the OECS*. <http://www.bb.undp.org/content/barbados/en/home/countryinfo/>

United Nations Statistics Division (UN). 2011. Per capita GDP at current prices - US dollars, *A data access system to UN databases (UNdata)*. <http://data.un.org/Default.aspx>

United Nations University (UNU). 2014. *World Income Inequality Database*, Version 3.0B. http://www.wider.unu.edu/research/WIID3-0B/en_GB/database/

Wikipedia. 2014. GPI Gini, *List of countries by income equality*. http://en.wikipedia.org/wiki/List_of_countries_by_income_equality

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*. [18](http://www.who.int/entity/healthinfo/global_burden_</p>
</div>
<div data-bbox=)

Appendix 1 Base data used in this paper

No.	Language name	Language code	No. of countries	Population (10**3)	S-rate log10(S)	H-rate log10(H)	GDP log10(gdp per capita)	GINI (Gini coefficient)	Fea-1 (Feature 2)	Fea-2 (Feature 6)	TMP Average annual temperature (degree C)	PRC Average annual precipitation (cm)
1	Albanian	alb	1	3014.579	0.880812	0.821314	3.393771	31.09	-1	-1	15.2	145.9
2	Amharic	amh	1	74066.147	0.771504	1.286554	2.128010	29.86	1	1	16.6	114.5
3	Arabic (Egyptian)	aeg	1	70591.288	0.200180	0.105408	3.048024	32.26	-1	-1	21.7	3.5
4	Arabic (Gulf)	arg	4	168091.237	1.061662	0.847333	3.363168	33.06	-1	-1	25.8	207.5
5	Arabic (Iraqi)	arq	1	26673.536	1.184603	0.865917	3.137731	28.60	-1	-1	22.7	15.5
6	Arabic (Modern Standard)	ams	11	115864.026	0.684713	1.063120	3.557097	35.77	-1	-1	23.7	40.3
7	Arabic (Moroccan)	amr	2	32911.245	0.359943	0.380008	3.246241	40.81	-1	-1	18.3	47.7
8	Arabic (Syrian)	asy	1	17676.012	-0.279631	0.412644	3.152063	35.78	-1	-1	16.9	17.6
10	Azerbaijani	aze	1	8306.500	0.108943	0.451343	3.019127	16.23	1	1	14.2	36.2
12	Bulgarian	bul	1	7781.161	1.136833	0.478457	3.511788	28.73	-1	1	10.3	58.6
13	Burmese	brm	1	49875.169	0.996099	1.196518	2.579350	38.10	1	-1	27.4	210.8
14	Catalan	ctl	1	79.060	0.848889	-0.079673	4.467936	28.10	-1	-1	11.0	107.2
15	Czech	cze	1	10197.101	1.230210	0.126426	4.048339	26.82	-1	1	8.4	49.9
16	Danish	dsh	1	5404.523	1.155397	-0.019535	4.655926	24.60	-1	1	9.1	58.2
19	English	eng	42	706106.354	0.937789	1.036576	4.367388	40.75	-1	1	18.3	120.7
20	Estonian	est	1	1362.550	1.399833	0.949131	3.945964	34.27	-1	1	6.0	67.9
21	Finnish	fin	1	5228.172	1.319829	0.407808	4.558260	28.27	-1	1	5.3	67.9
22	French	fre	18	249364.700	0.997206	1.269611	3.934029	40.37	-1	-1	22.0	117.9
23	Georgian	geo	1	4318.300	0.287105	0.564820	3.074404	39.78	1	1	12.7	53.1
25	Greek (Modern)	grk	2	12071.549	0.469537	-0.086086	4.315603	33.49	-1	1	18.9	37.1
26	Hebrew (Modern)	heb	1	6809.000	0.787081	0.676414	4.269251	41.20	-1	-1	15.9	63.2
27	Hindi	hin	1	1110626.108	1.227319	0.738909	2.812720	33.15	1	1	25.2	76.8
29	Icelandic	ice	1	292.074	1.076481	0.005114	4.656770	27.81	-1	1	4.7	84.7
30	Indonesian	ind	1	221293.797	1.031211	0.954456	3.064688	32.59	-1	-1	28.0	182.1
31	Irish	iri	1	4070.262	1.092280	-0.138096	4.660547	33.79	-1	-1	9.8	77.5
32	Italian	ita	2	57714.617	0.853405	-0.018780	4.478474	37.26	-1	-1	15.4	70.7
33	Japanese	jpn	1	127761.000	1.395182	-0.270014	4.561596	32.11	1	1	16.3	152.9
34	Khalkha	kha	1	2496.248	1.059673	0.511418	2.902016	34.07	1	1	-0.1	28.1
35	Khmer	khm	1	13149.386	0.651690	1.266511	2.608460	35.53	-1	-1	27.8	140.7
36	Kinyarwanda	kin	1	9254.379	0.887160	1.424369	2.353630	52.56	-1	-1	20.5	106.0
37	Korean	kor	2	71678.718	1.334424	0.891730	4.034854	31.44	1	1	11.9	130.4
38	Lao	lao	1	5699.112	1.311896	0.729125	2.618281	33.67	-1	-1	26.7	168.2
39	Latvian	lat	1	2263.122	1.418599	1.008700	3.783960	34.96	-1	1	6.0	70.2
40	Lithuanian	lit	1	3377.075	1.632716	0.959552	3.826706	35.20	-1	1	7.1	63.1
41	Macedonian	mcd	1	2085.728	0.939178	0.717185	3.422229	38.85	-1	1	12.0	57.1
42	Mandarin	mmd	2	1300241.700	1.228281	0.323744	3.196846	42.53	-1	1	12.9	53.9
43	Motu	mtu	1	5948.461	0.988738	1.180714	2.819669	50.88	1	-1	26.8	323.6
44	Nauruan	nau	1	10.092	0.941204	1.077837	3.725319	66.86	-1	-1	27.8	205.1
45	Nepali	nep	1	24921.910	0.994556	1.134217	2.465188	42.26	1	1	19.0	147.6
46	Norwegian	nor	1	4591.910	1.064093	-0.081523	4.753029	30.17	-1	1	4.8	85.0
47	Pashto	psh	1	24018.682	0.796098	0.528509	2.342534	27.80	1	1	13.1	28.9
48	Persian	prs	1	69342.126	0.785484	0.404208	3.371794	39.11	1	-1	18.4	24.1
49	Polish	pol	1	38182.222	1.229780	0.207498	3.820863	35.43	-1	1	8.4	53.0
50	Portuguese	por	7	232924.896	0.779586	1.466058	3.575315	54.66	-1	-1	21.3	131.5
51	Romanian	rom	1	21451.748	1.105648	0.512864	3.548176	30.04	-1	-1	10.8	60.0
53	Russian	rus	1	143821.212	1.562512	1.472187	3.613777	40.77	-1	1	5.8	70.7
54	Samoan	sam	1	178.794	0.563156	0.053785	3.321107	46.29	-1	-1	26.5	271.1
55	Serbian-Croatian	scr	1	614.670	1.277682	0.431069	3.528011	30.14	-1	1	20.5	1.3
56	Sesotho	ses	1	1912.022	0.717011	1.122134	2.809892	51.92	-1	-1	15.5	65.0
57	Sinhala	snh	1	19435.000	1.450317	0.838138	3.026599	40.74	1	1	27.7	232.2
58	Slovene	slo	1	1997.012	1.450533	0.309473	4.229021	31.15	-1	1	10.7	140.1
59	Somali	som	1	8249.965	1.398209	0.521889	2.161562	39.70	1	-1	22.5	50.0
60	Spanish	spa	20	393990.340	0.806887	1.296117	3.821655	48.00	-1	-1	17.8	93.1
61	Swahili	swa	1	37765.139	0.882318	1.416664	2.543597	36.31	-1	-1	26.0	107.2
62	Swedish	swe	1	8993.531	1.138953	0.091901	4.604886	26.36	-1	1	6.7	53.6
64	Tajik	taj	1	6663.929	0.332681	0.343362	2.493528	33.60	1	-1	15.5	56.8
65	Thai	tha	1	65087.400	1.009878	0.848019	3.394245	42.16	-1	-1	28.9	165.3
66	Tigrinya	tig	1	4665.522	0.694181	1.201636	2.376052	40.00	1	1	16.0	35.4
67	Tongan	tng	1	100.319	0.516362	-0.196905	3.380264	37.71	-1	-1	23.5	163.8
68	Turkish	tur	1	66845.635	0.561978	0.463909	3.768397	41.29	1	1	12.1	40.8
69	Turkmen	tkm	1	4696.876	1.034130	0.946401	3.163142	40.77	1	1	16.8	20.7
70	Ukrainian	ukr	1	47451.600	1.429565	1.077609	3.135880	28.93	-1	1	8.4	61.1
71	Urdu	urd	1	155151.394	1.012720	0.550765	2.800372	30.92	1	1	23.0	47.0
72	Uzbek	uzb	1	25864.350	0.774537	0.545823	2.667565	35.19	1	1	14.8	45.1
73	Vietnamese	vie	1	81437.700	0.999000	0.576957	2.783113	36.81	-1	-1	24.5	164.5