Relation between Word Order Characteristics and Suicide/Homicide Rates (2)
語順特徴と自殺率／他殺率との関係(その 2)

Terumasa EHARA
江原暉将

Yamanashi Eiwa College
山梨英和大学
http://www.y-eiwa.ac.jp/~kenkyu/eharate/

1 Introduction

The previous paper [Ehara, 2010] shows quantitative relations between the word order characteristics and suicide/homicide rates. Our study purpose is to clarify relations between syntactic structures, especially word order, of a language and people's thinking pattern who speak that language [Ehara, 1995], [Ehara, 2007]. In this paper, we expand the experimental conditions of the previous paper.

Death is the most important event for all human beings. Suicide and homicide are abnormal death. Then, we think that people's thinking pattern affects the suicide and homicide. We can measure them quantitatively by suicide rate and homicide rate. So, we use them as the measure of thinking pattern.

2 Data

Data for the word order characteristics (features) are obtained from the WALS database [Dryer, 2005]. The number of languages analyzed in this article is 1473. The following thirteen word order features are considered in our paper.

(1) Order of Subject(S) and Verb(V)
(2) Order of Object(O) and Verb
(3) Order of Oblique(X) and Verb
(4) Order of Adposition(Ad) and Noun Phrase(N)
(5) Order of Genitive(G) and Noun
(6) Order of Adjective(A) and Noun
(7) Order of Demonstrative(Dm) and Noun
(8) Order of Numeral(Nm) and Noun
(9) Order of Relative Clause(R) and Noun
(10) Order of Degree Word(Dg) and Adjective
(11) Position of Polar Question Particles
(12) Position of Interrogative Phrases in Content Questions
(13) Order of Adverbial Subordinator(As) and Clause(C)

Feature value is "+" if the order is same as in Japanese and "－" if the order is opposite of the "+". The feature value is "0" if the order is other than "+" or "－" and "." if the feature value is not described in WALS database.

For example, as Japanese has OV order for the feature number 2, OV order is represented as "+" and VO order is represented as "－". "No dominant order" is represented as "0". Table 1 shows all feature values "+" and "－" for the thirteen features.

Table 1: Word order feature values

No.	+	－
1	SV	VS
2	OV	VO
3	XV	VX
4	NAp	ApN
5	GN	NG
6	AN	NA
7	DmN	NDm
8	NmN	NNm
9	RN	NR
10	DgA	ADg
11	Final	Initial
12	Not initial	Initial
13	CAs	AsC

Suicide rate and homicide rate[1] are obtained from the WHO's "mortality and burden of disease estimates for WHO member states in 2004" [WHO, 2009]. From it, we can get "the number of death by suicide" and "the

[1] Suicide rate is defined as follows. If Ns is the number of death by suicide in one year and N is the population, then suicide rate is Ns/N multiplied by 100 thousands. Homicide rate is defined same as above. The number of death by homicide excludes the death by war.

number of death by homicide". In this database, "suicide" is represented as "self-inflicted injuries" (GBD code: W157) and "homicide" is represented as "violence" (GBD code: W158). "Violence" does not include "war" (GBD code: W159). From this database, we can obtain suicide and homicide rate for 192 countries or regions of the world.

Language names spoken in countries and regions are obtained from Nations Online [Nationsonline, 2006]. This table includes 218 country names with official or national language names. We use the firstly listed language name in the table as the language name spoken in the country.

Combining the above three databases, we get 177 country names and 67 language names[2]. Our previous paper, only, treats 98 countries and 34 languages. So, we can extend our experimental conditions.

Next, we merge the countries speaking a same language. For example, 42 English speaking countries are merged and 20 Spanish speaking countries are, also, merged. In this merging process, the numbers of death by suicide and by homicide are summed up. As the result, we can obtain the table shown in Appendix 1, of which each item includes language name, number of death by suicide, number of death by homicide, thirteen word order feature's values. This table includes 67 items.

3 Analysis and results

We define suicide / homicide ratio (SH-ratio) as :

$$SH-ratio = \log_{10}\left(\frac{suicide\ \ rate}{homicide\ \ rate}\right)$$

$$= \log_{10}\left(\frac{the\ number\ of\ \ death\ by\ suicide}{the\ number\ of\ \ death\ by\ homicide}\right)$$

From this definition, we need not know population of countries. In this computation, if either of "the number of death by suicide" or "the number of death by homicide" is zero, SH-ratio cannot be computed. And SH-ratio is less reliable in the case of either of the numbers is low. So, if either of the numbers is less than 10, we omit that item. Computed SH-ratios are shown in Appendix 1.

Next, we make t-tests for the thirteen features. For each features, we divide SH-ratio data to the plus group and the minus group and compute the t-value. The result is shown in Table 2.

Table 2 T-test results for the thirteen features

feature number	data counts		sample mean		sample standard deviation		t-value
	+	−	+	−	+	−	
1	55	4	0.298	-0.070	0.564	0.841	2.008
2	20	41	0.199	0.283	0.509	0.610	-1.871
3	3	17	0.890	0.249	0.641	0.572	2.221
4	17	42	0.331	0.333	0.538	0.585	-0.046
5	25	30	0.354	0.146	0.554	0.593	4.750
6	38	27	0.519	0.036	0.522	0.502	14.385
7	47	10	0.387	0.111	0.569	0.525	3.799
8	51	6	0.426	-0.190	0.558	0.434	5.477
9	9	43	0.363	0.334	0.604	0.584	0.344
10	36	6	0.513	-0.030	0.510	0.385	5.080
11	13	18	0.156	0.251	0.644	0.491	-1.187
12	27	20	0.279	0.437	0.539	0.670	-2.909
13	5	47	0.402	0.358	0.743	0.561	0.301

The largest t-value is 14.385 observed at the feature 6: "order of adjective and noun". Figure 1 shows histograms of SH-ratios for the "+"(AN) group and the "−"(NA) group. The Kolmogorov-Smirnov test for normality shows both these two distributions are normally distributed with the confidence level 5%. T-value 14.385 for the degree of freedom 63 means two mean values are differ with the confidence level 1%.

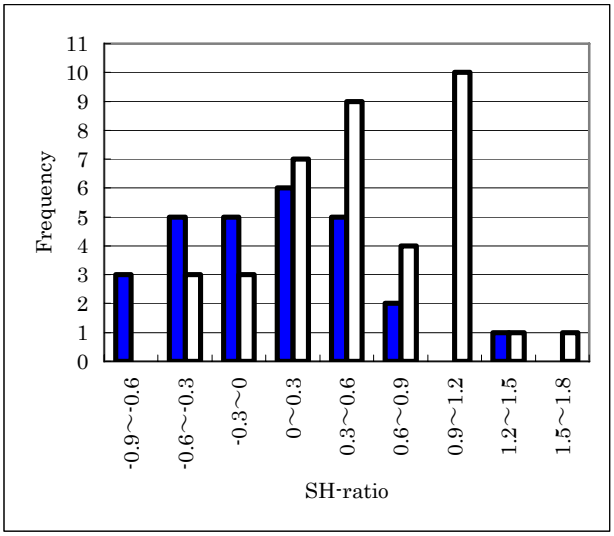
4 Conclusion

We examine the relation between word order characteristics and suicide / homicide rates. From the results, we can conclude that there is relation between SH-ratio and Adjective and Noun word order. It is from the t-test of "adjective noun group" and "noun adjecive group".

Economical and political situation of countries may affect SH-ratio. Climate condition may, also, affect the ratio. Linguistic features

[2] In the combination of databases, we make some normalization for the country names and the language names. For example, "United States of America" and "United States" are normalized to the former. "Chinese" and "Mandarin" is normalized to the latter.

which are not used in this paper may affect the ratio. Study using these features is remained for the future work.



Dark bar: NA language group

Light bar: AN language group

Figure 3　Histograms of SH-ratios

5 References

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

[Ehara, 1995] EHARA, Terumasa : Relation among Word Order Parameters Analyzed by Multi-Dimensional Scaling, Proceedings of The first Annual Meeting of The Association for Natural Language Processing, pp.173-176, Mar., 1995 (in Japanese).

[Ehara, 2007] EHARA, Terumasa : Word Order Characteristics Analyzed by Multi Dimensional Scaling, Proceedings of The 13th Annual Meeting of The Association for Natural Language Processing, A1-3, Mar., 2007.

[Ehara, 2010] EHARA, Terumasa : 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, Mar., 2010.

[Nationsonline, 2006] One World - Nations Online : Official and National Languages of the World by Continent, 2006 version. http://www.nationsonline.org/oneworld/languages.htm

[WHO, 2009] World Health Organization : Mortality and burden of disease estimates for WHO member states in 2004. http://www.who.int/entity/healthinfo/global_buden_disease/gbddeathdalycountryestimates2004.xls

Appendix 1　Data used in the analysis

language name	number of death by suicide	number of death by homicide	SH-ratio	feature_1	feature_2	feature_3	feature_4	feature_5	feature_6	feature_7	feature_8	feature_9	feature_10	feature_11	feature_12	feature_13
Tagalog	1395	17486	-1.10	−	−	.	.	−	0	0	+	−	.	0	−	−
Arabic (Syrian)	97	476	-0.69	0	−	.	−	−	−	0	0	−	0	+	−	−
Portuguese	13989	68035	-0.69	+	−	.	−	−	−	+	.	−	+	−	.	.
Khmer	615	2534	-0.61	+	−	.	−	−	−	−	−	−	−	+	+	−
Rundi	698	2677	-0.58	+	.	.	.	.	.	.	.	.	.	.	.	.
Kinyarwanda	698	2405	-0.54	+	−	−	−	.	−	+	−	.	.	.	+	.
Swahili	2861	9790	-0.53	+	−	.	−	−	−	−	−	−	.	−	+	−
Amharic	4550	14894	-0.51	+	+	.	+	+	+	+	+	+	.	+	.	+
Tigrinya	215	693	-0.51	+	+	.	.	+	+	+	.	.	.	.	.	.
Spanish	25109	78844	-0.50	0	−	−	−	−	−	+	+	−	+	0	−	−
Sesotho	102	260	-0.41	+	−	−	−	−	−	−	.	−	.	0	.	.
Arabic (Modern Standard)	6060	14885	-0.39	−	−	−	−	−	−	+	+	.	−	−	.	−
Azerbaijani	107	235	-0.34	+	+	.	.	+	+	.	+	.	.	+	+	.
Georgian	87	166	-0.28	+	+	.	+	+	+	+	+	.	.	.	+	−
French	22427	39251	-0.24	+	−	−	−	−	−	+	+	−	+	−	−	−
Burmese	4714	7478	-0.20	+	+	0	+	+	−	+	−	+	+	0	+	0
Motu	578	900	-0.19	+	+	.	+	+	−	.	.	.	.	+	+	.

language name	number of death by suicide	number of death by homicide	SH-ratio	feature_1	feature_2	feature_3	feature_4	feature_5	feature_6	feature_7	feature_8	feature_9	feature_10	feature_11	feature_12	feature_13
Nepali	2622	3617	-0.14	+	+	.	+	+	+	+	+	.	.	.	+	+
English	61877	77942	-0.10	+	−	−	−	0	+	+	+	−	+	0	−	−
Arabic (Moroccan)	751	766	-0.01	−	−	.	−	−	−	+	+	−	−	−	.	−
Tajik	139	143	-0.01	+	+	.	−	−	−	+	+	−	.	+	.	−
Albanian	238	208	0.06	+	−	.	−	−	−	+	+	−	+	−	.	−
Indonesian	23986	20100	0.08	+	−	−	−	−	−	−	+	−	+	0	0	−
Armenian (Eastern)	121	100	0.08	+	.	.	+	+	+	+	+	.	.	.	.	−
Arabic (Egyptian)	1134	912	0.09	+	−	−	−	−	−	−	0	−	−	−	+	−
Russian	52841	42918	0.09	+	−	.	−	−	+	+	+	−	+	0	−	−
Turkmen	516	421	0.09	+	+	.	+	+	+	+	+	+	.	.	.	.
Turkish	2627	2096	0.10	+	+	+	+	+	+	+	+	+	+	+	+	0
Hebrew (Modern)	403	312	0.11	+	−	.	−	−	−	−	+	−	0	−	−	−
Thai	6400	4409	0.16	+	−	−	−	−	−	−	−	−	−	+	+	−
Arabic (Gulf)	20496	12531	0.21	+	−	−	−	−	−	0	+	−	+	0	+	−
Macedonian	176	106	0.22	+	−	.	−	0	+	+	.	.	.	.	.	.
Uzbek	1559	921	0.23	+	+	.	.	+	+	+	+	+	+	0	+	−
Pashto	1506	813	0.27	+	+	.	0	+	+	+	+	−	+	.	+	−
Arabic (Iraqi)	4200	2016	0.32	+	−	.	−	−	−	.	+	−	0	.	+	−
Ukrainian	12713	5653	0.35	+	−	.	−	.	+	+	+	−	.	−	.	−
Persian	4190	1742	0.38	+	+	.	−	−	−	+	+	−	+	−	+	−
Latvian	607	236	0.41	0	−	.	−	+	+	+	+	−	+	−	−	−
Vietnamese	8365	3165	0.42	+	−	−	−	−	−	−	+	−	.	+	+	−
Korean	15369	5555	0.44	+	+	.	+	+	+	+	+	+	+	0	+	+
Estonian	339	120	0.45	+	−	−	+	+	+	+	+	−	+	0	−	−
Urdu	15995	5521	0.46	+	+	.	+	+	+	+	+	0	+	−	+	−
Hindi	188524	61229	0.49	+	+	.	+	+	+	+	+	0	+	−	+	−
Belorussian	3535	1007	0.55	0	0	0	.	−	+	+	.	−	.	−	.	.
Khalkha	293	83	0.55	+	+	.	+	+	+	+	+	+	+	+	+	+
Greek (Modern)	356	99	0.56	0	−	.	−	−	+	+	+	−	+	−	−	−
Lao	1143	299	0.58	+	−	.	−	−	−	−	−	−	−	+	+	−
Romanian	2771	708	0.59	+	−	.	−	−	−	0	+	−	+	0	−	−
Sinhala	5370	1312	0.61	+	+	.	+	.	+	.	0	−	+	.	.	0
Bulgarian	1068	235	0.66	0	−	−	−	0	+	+	+	−	+	0	−	−
Lithuanian	1477	313	0.67	+	−	.	−	+	+	+	+	−	+	−	−	−
Serbian	1993	284	0.85	+	−	.	−	0	+	+	+	.	+	0	.	−
Italian	4174	559	0.87	0	−	.	−	−	−	+	+	−	+	0	.	−
Somali	1990	265	0.88	+	+	.	0	0	−	0	+	−	+	0	+	−
Mandarin	222722	27748	0.90	+	−	+	−	+	+	+	+	+	+	+	+	.
Finnish	1093	134	0.91	+	−	−	+	+	+	+	+	−	+	0	−	−
Dutch	3750	420	0.95	+	0	0	−	−	+	+	+	−	.	0	.	−
Polish	6492	617	1.02	0	−	.	−	−	+	+	+	−	+	−	−	−
Swedish	1239	111	1.05	+	−	−	−	+	+	+	+	−	+	0	−	−
Hungarian	2727	222	1.09	+	0	0	+	+	+	+	+	0	+	.	+	−
Czech	1732	136	1.11	+	−	.	−	0	+	+	+	−	.	0	+	−
Slovene	564	41	1.14	+	−	.	−	0	+	+	+	.	.	.	.	.
Norwegian	534	38	1.15	+	−	.	−	0	+	+	+	−	+	0	−	−
Danish	773	52	1.17	+	−	−	−	+	+	+	+	−	+	0	−	−
Irish	503	30	1.22	−	−	−	−	−	−	−	+	−	.	−	−	−
German	14428	701	1.31	+	0	0	−	−	+	+	+	−	+	0	−	−
Japanese	31747	686	1.67	+	+	+	+	+	+	+	+	+	+	+	+	+