Relation between the word order characteristics and suicide/homicide rates
語順特徴と自殺率／他殺率との関係

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

There are a lot of studies for the relation between language structures and thinking patterns. However, many of them are non quantitative study and are concerning for lexical or morphological structures of languages. From the author's knowledge, there are few quantitative studies for the relation between human thinking patterns and syntactic structures. [Kashima, 1998] [Kashima, 2003] make quantitative studies for the relation between culture types and "pronoun drop".

Word order is one of the major syntactic structures investigated in the language typology field. [Ehara, 1995] makes short glance at the relation between human thinking pattern and the word order. On the other hand, [Haga, 2004] makes a negative description for this relation. However, this description is not quantitative.

This article presents quantitative study results for the relation between the word order characteristics and the suicide / homicide rates. Why do we use suicide / homicide rates? Death is the most important event for all human beings. Suicide and homicide are abnormal death. Then they are, I suppose, affected by the people's thinking pattern. And we can measure them quantitatively by suicide rate and homicide rate. Then we use them as the measure of thinking pattern.

2 Data

Data for the word order characteristics are obtained from [Ehara, 2007]. The number of languages analyzed in that article is 576. The following seven word order characteristics are considered in it.

(1) Subject(S) and Verb(V) in a declarative sentence.
(2) Object(O) and Verb(V) in a declarative sentence
(3) Noun(N) and Adposition(Ap)
(4) Genitive(G) and Noun(N)
(5) Adjective(A) and Noun(N)
(6) Demonstrative(Dm) and Noun(N)
(7) Numeral(Nm) and Noun(N)

Suicide rate and homicide rate[1] are obtained from the WHO's mortality database [WHO, 2006]. From it, we can get "the number of death by suicide" and "the number of death by homicide". Then we can obtain suicide and homicide rate for 151 countries (or regions) of the world.

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

Combining these three databases, we get

[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 wars.

98 country names and 34 language names[2]. Next, we merge the countries speaking a same language. For example, 24 English speaking countries are merged and 21 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, seven word order characteristic's values, value of the first principal component and value of the second principal component. The last two data columns are obtained from the multi dimensional analysis for the word order characteristics described in [Ehara, 2007]. This table includes 34 items.

3 Analysis and results

Firstly, 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, we add one to the numbers of death by suicide and by homicide when they are zero. SH ratio are shown in Appendix 1.

Next, we compute Pearson's correlation coefficient between SH ratio and the first principal component value and between SH ratio and the second principal component value for 34 languages. Figure 1 shows the former relation and Figure 2 shows the latter relation. The former correlation coefficient is −0.0699 that means they don't have correlation at a significant level 5%. The latter correlation coefficient is −0.5748 that means they have correlation at a significant level 1%.

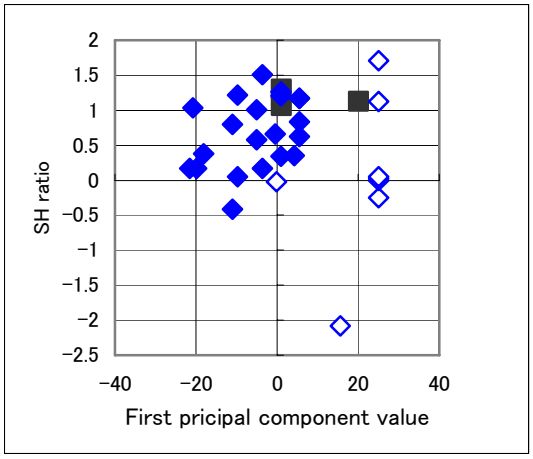

Figure 1 Relation between SH ratio and first principal component value

dark diamond : 23 VO languages
light diamond : 8 OV languages
square : 3 no dominant OV or VO order languages

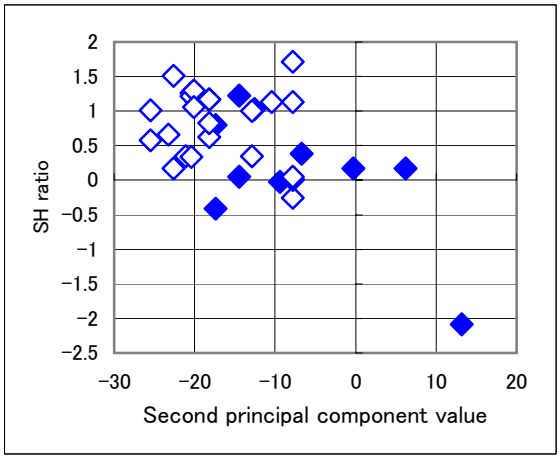

Figure 2 Relation between SH ratio and second principal component value

dark diamond : 10 NA languages
light diamond : 24 AN languages

The data point at the bottom-right position in Figure 2 is Burmese point. Eliminating

[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.

this point, the correlation between SH ratio and second principal component value is −0.3421 that, also, means they have correlation at a significant level 1%.

Next, we make t-tests for seven features. For each features, we divide SH ratio data to a plus group and a minus group and compute t-value. The result is shown in Table 1.

Table 1 T-test results for seven features

Feature No.		1	2	3	4	5	6	7
Number of data	+	27	8	10	15	24	30	31
	−	1	23	24	15	10	4	2
Sample mean	+	0.58	0.07	0.31	0.48	0.77	0.60	0.69
	−	1.03	0.68	0.69	0.60	0.13	0.44	−0.96
Sample variance	+	0.58	1.07	1.01	0.76	0.28	0.55	0.30
	−	0.00	0.23	0.25	0.31	0.77	0.12	1.27
t-value		----	−4.96	−3.68	−1.16	6.41	0.68	3.58

The largest t-value is observed at feature 5 i.e. AN and NA order. Figure 3 shows histograms of SH ratios for the AN group and the NA group. They may be non-normal distributions. So these t-values are only a guide.

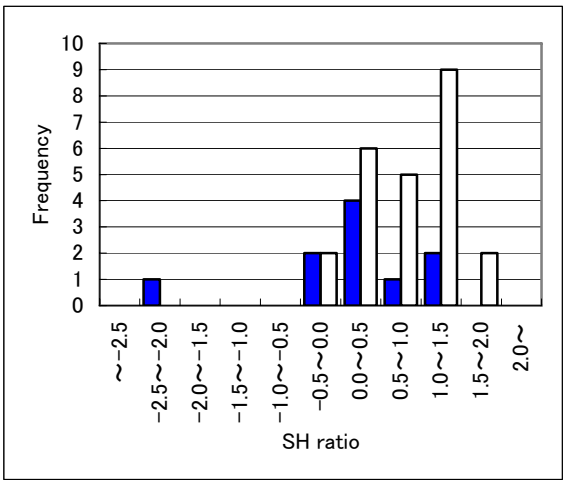

Figure 3 Histograms of SH ratios for the NA language group (dark bar) and the AN language group (light bar)

4 Conclusion

We examine the relation between word order characteristics and suicide / homicide rates. From the results, we can recognize a correlation between suicide / homicide ratio (SH ratio) and the second principal component value obtained by multi dimensional analysis for word order characteristics. We can, also, observe that there is relation between SH ratio and Adjective and Noun word order by t-test.

Economical and political situation of countries may affect SH ratio. Climate condition may affect the ratio. Linguistic features which are not used in this article may affect the ratio. Study using exhaustive features is remained for the future works. Increasing the number of countries and languages are also the future works.

5 References

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

[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.

[Haga, 2004] 芳賀綏：日本人らしさの構造－言語文化論講義－, 大修館書店, pp.164-167, 2004.

[Kashima, 1998] Kashima, Emiko S. and Kashima, Yoshihisa：Culture and Language - The Case of Cultural Dimensions and Personal Pronoun Use -, Journal of Cross-Cultural Psychology, Vol. 29, No. 3, 461-486, 1998.

[Kashima, 2003] Kashima, Yoshihisa and Kashima, Emiko S.：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, 125-134, 2003.

[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, 2006] World Health Organization : WHO Mortality Database, 2006 version. http://www.who.int/whosis/mort/en/

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	First pricipal component value	Second pricipal component value
Burmese	3	361	−2.08	+	+	+	+	−	+	−	15.6	13.2
Spanish	21151	54614	−0.41	0	−	−	−	−	+	+	−11.1	−17.4
Georgian	100	178	−0.25	+	+	+	+	+	+	+	25.0	−7.8
Persian	234	246	−0.02	+	+	−	−	−	+	+	−0.1	−9.4
Urdu	1	1	0.00	+	+	+	+	+	+	+	25.0	−7.8
Turkish	144	137	0.02	+	+	+	+	+	+	+	25.0	−7.8
Turkmen	241	216	0.05	+	+	+	+	+	+	+	25.0	−7.8
Albanian	146	131	0.05	+	−	−	−	−	+	+	−9.7	−14.5
Russian	42855	28844	0.17	+	−	−	−	+	+	+	−3.7	−22.6
Arabic	121	81	0.17	+	−	−	−	−	−	0	−19.8	−0.3
Thai	4905	3332	0.17	+	−	−	−	−	−	−	−21.5	6.2
English	44213	20141	0.34	+	−	−	0	+	+	+	0.9	−20.4
Estonian	273	123	0.35	+	−	+	+	+	+	+	15.4	−12.9
Latvian	487	218	0.35	0	−	−	+	+	+	+	4.2	−21.1
Hebrew	659	273	0.38	+	−	−	−	−	−	+	−18.1	−6.7
Greek	413	109	0.58	0	−	−	−	+	+	+	−5.1	−25.5
Lithuanian	1049	254	0.62	+	−	−	+	+	+	+	5.5	−18.2
Bulgarian	1013	222	0.66	0	−	−	0	+	+	+	−0.4	−23.3
Italian	4076	649	0.80	0	−	−	−	−	+	+	−11.1	−17.4
Mandarin	31852	4715	0.83	+	−	−	+	+	+	+	5.5	−18.2
Finnish	1061	105	1.00	+	−	+	+	+	+	+	15.4	−12.9
Polish	5805	569	1.01	0	−	−	−	+	+	+	−5.1	−25.5
Irish	392	37	1.03	−	−	−	−	−	−	+	−20.8	−12.6
Dutch	3733	327	1.06	+	0	−	−	+	+	+	1.1	−20.1
Korean	10688	790	1.13	+	+	+	+	+	+	+	25.0	−7.8
Hungarian	2621	194	1.13	+	0	+	+	+	+	+	20.2	−10.4
Swedish	1219	82	1.17	+	−	−	+	+	+	+	5.5	−18.2
Danish	727	49	1.17	+	−	−	+	+	+	+	5.5	−18.2
Czech	1564	96	1.21	+	−	−	0	+	+	+	0.9	−20.4
French	10812	649	1.22	+	−	−	−	−	+	+	−9.7	−14.5
Norwegian	533	29	1.26	+	−	−	0	+	+	+	0.9	−20.4
German	12357	614	1.30	+	0	−	−	+	+	+	1.1	−20.1
Icelandic	32	1	1.51	+	−	−	−	+	+	+	−3.7	−22.6
Japanese	29921	580	1.71	+	+	+	+	+	+	+	25.0	−7.8

Feature values :

+　: SV, OV, N Ad, GN, AN, Dm N, Nm N order for first to seventh features respectively.

−　: VS, VO, Ad N, NG, NA, N Dm, N Nm order for first to seventh features respectively.

0　: no dominant order.