Oblique and verb word order relates to speakers' thought patterns

斜格要素と動詞の語順と話者の思考パターンとの関係

Terumasa EHARA
江原暉将

Ehara NLP Research Laboratory
江原自然言語処理研究室

eharate @ gmail.com

Abstract

The author has quantitatively investigated the relationship between the word order of languages and speakers' thought patterns. In this paper, it will be focused on oblique (X) and verb (V) word order feature (X:V). There are not many languages for which the word order feature values for X:V was revealed. So, I first expand the feature values for X:V using the Universal Dependency Database. After expansion, correlation coefficients between nine word order features including X:V and four quantitative metrics reflecting thought patterns (thought pattern metrics) are calculated. Results show that X:V has significant correlations with all four thought pattern metrics, although object (O) and verb (V) word order feature (O:V), which is considered the most important word order feature, has significant correlations with only two thought pattern metrics.

1 Introduction

The author has quantitatively investigated the relationship between the word order of languages and speakers' thought patterns. Word order (feature) means the order of dependent and head in the dependency grammar. Word order feature values are head initial or head final (sometimes no dominant order). Although I have focused on the word order of object (O) and verb (V) (O:V), which is considered the most important word order feature, I focus on the word order of oblique (X) and verb (V) (X:V) in this paper. The reasons why oblique is focused on are the following: "Place" which is emphasized in the linguistics of place [1] is expressed by oblique elements in the sentences. The "Topic" element in the topic prominent languages [2] (other than subject or object) is oblique.

There are not many languages for which the word order feature value for X:V was revealed. For example, in WALS [3], O:V feature values (OV, VO, or No dominant order (NDO)) are described for 1,519 languages out of 2,679 languages (57%). On the other hand, X:V feature values (XV, VX, or NDO) are described only for 499 languages (19%). In section 2, I will expand the X:V feature values using the Universal Dependency Database [4].

Four quantitative metrics reflecting thought patterns, which have been used in the previous papers, are also used in this paper. They are the following: The ratio of suicide rate and homicide rate (S/H_rate)[i] [5], the entropy of the distribution of Y chromosome DNA haplogroup (Y_DNA) [6], the entropy of the distribution of Mitochondrial DNA sub-haplogroup (Mt_DNA) [7] and the ratio of S allele in serotonin transponder gene (S_ratio) [8]. These metrics are called thought pattern metrics. Details of these metrics should be referred to the references. In section 3, we examine the relation between the four thought pattern metrics and the nine word order features including X:V.

[i] $\mathrm{S/H\_rate} = \log_{10} \mathrm{S\_rate} - \log_{10} \mathrm{H\_rate}$, where S_rate is the suicide rate and H_rate is the homicide rate.

2 Expansion of X:V feature values

We expand X:V feature values using the Universal Dependency Database[ii] (UDD) [4]. UDD describes the dependency relations for 138 languages and the dependency relation between oblique and verb is described as the "obl" relation.

Table 1 Comparison of r value obtained from the UDD and WALS feature value

UDD_name	WALS_name	r	x:v_wals
Coptic	Coptic	0.04	VX
Bambara	Bambara	0.06	VX
Irish	Irish	0.08	VX
Scottish_Gaelic	Gaelic (Scots)	0.08	VX
Arabic	Arabic (Modern Standard)	0.11	VX
English	English	0.12	VX
Thai	Thai	0.13	VX
Wolof	Wolof	0.15	VX
French	French	0.16	VX
Yupik	Yupik (Siberian)	0.17	VX
Danish	Danish	0.18	VX
Swedish	Swedish	0.18	VX
Indonesian	Indonesian	0.19	VX
Welsh	Welsh	0.21	VX
Spanish	Spanish	0.22	VX
Karo	Karó (Arára)	0.24	VX
Guajajara	Guajajara	0.25	VX
Vietnamese	Vietnamese	0.25	VX
Bulgarian	Bulgarian	0.33	VX
Belarusian	Belorussian	0.36	NDO
Finnish	Finnish	0.38	VX
Guarani	Guaraní	0.39	VX
Kaapor	Urubú-Kaapor	0.4	NDO
Frisian_Dutch	Frisian	0.41	NDO
Mbya_Guarani	Guaraní	0.48	VX
Apurina	Apurinã	0.5	NDO
Estonian	Estonian	0.51	VX
Dutch	Dutch	0.53	NDO
Komi_Zyrian	Komi-Zyrian	0.55	VX
Akuntsu	Mekens	0.59	VX
Western_Armenian	Armenian (Western)	0.65	NDO
Armenian	Armenian (Eastern)	0.66	NDO
Cantonese	Cantonese	0.69	XV
German	German	0.7	NDO
Warlpiri	Warlpiri	0.71	NDO
Basque	Basque	0.73	XV
Chinese	Mandarin	0.89	XV
Bhojpuri	Bhojpuri	0.98	XV
Turkish	Turkish	0.98	XV
Bengali	Bengali	1	XV
Japanese	Japanese	1	XV
Tamil	Tamil	1	XV

In obl relations, I select the relations in which the part of speech of the dependent (X) is the noun (NOUN), pronoun (PNON) or proper noun (PROPN), and the part of speech of the head (V) is the verb (VERB), adjective (ADJ) or adverb (ADV), as the X:V relations. For data of each language, I define head final ratio $r = n_{XV}/(n_{XV} + n_{VX})$ where n_{XV} is the number of XV word order data and n_{VX} is the number of VX word order data. These values for 138 languages are shown in the Appendix.

Table 1 shows r values and X:V feature values for 42 languages for which X:V feature values are described in WALS and r values are available by the UDD. The X:V feature values in $r \leq 0.33$ are all VX in WALS, and X:V feature values in $0.73 \leq r$ are all XV. No conflict between WALS and UDD exists.

3 Relation between word order features and thought pattern metrics

I use nine word order features shown in Table 2 including X:V. Feature values are obtained mainly from WALS and several values are from Yamamoto [9]. Yamamoto does not describe the X:V feature values, so X:V feature values are complemented by the UDD. X:V feature values of languages that $r < 0.5$ are set as VX and X:V feature values of languages that $0.5 \leq r$ are set as XV. No feature values other than X:V are complemented by the UDD.

To calculate the correlation between word order features and thought pattern metrics, head initial feature values are quantified to -1, and head final feature values are quantified to 1. Other feature values (such as NDO) are not used (regarded as No Data).

Table 2 Word order features used in the research

Dependent	Head
Subject (S)	Verb (V)
Object (O)	Verb (V)
Oblique (X)	Verb (V)
Noun (N)	Adposition (P)
Genitive (G)	Noun (N)
Adjective (A)	Noun (N)
Demonstrative (D)	Noun (N)
Numeral (Nu)	Noun (N)
Relative clause (R)	Noun (N)

Table 3 shows Spearman's rank correlation coefficients, two-sided probabilities, and the number of data between nine word order feature values (after quantification) and

[ii] http://hdl.handle.net/11234/1-4923 (accessed on 24 Feb. 2023).

four thought pattern metric values (real values)[iii]. College Analysis Version 8.6 [iv] [10] is used to calculate these correlation coefficients.

Table 3 Correlation coefficients between word order features and thought pattern metrics

(a) Spearman's rank correlation coefficient

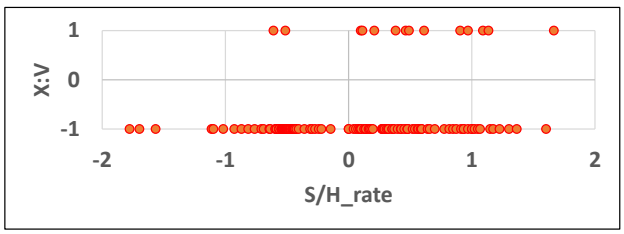
Metrics ＼ word order feature	S:V	O:V	X:V	N:P	G:N	A:N	D:N	Nu:N	R:N
S/H_rate	-0.005	0.114	0.213	0.159	0.262	0.266	0.070	-0.022	0.073
Y_DNA	0.185	0.208	0.246	0.109	0.127	0.306	0.174	0.277	0.306
Mt_DNA	0.220	0.262	0.788	0.288	0.405	0.450	0.586	0.435	0.355
S_ratio	0.008	0.463	0.662	0.399	0.480	0.330	0.686	0.237	0.476

(b) Two-sided Probability

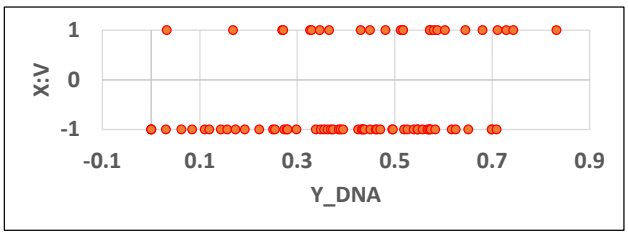
Metrics ＼ word order feature	S:V	O:V	X:V	N:P	G:N	A:N	D:N	Nu:N	R:N
S/H_rate	0.9519	0.1373	0.0102	0.0361	0.0029	0.0003	0.3832	0.7736	0.3395
Y_DNA	0.0181	0.0070	0.0270	0.1687	0.1062	0.0001	0.0367	0.0005	0.0001
Mt_DNA	0.3664	0.2269	0.0005	0.2185	0.0684	0.0310	0.0134	0.0432	0.1046
S_ratio	0.9170	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0028	0.0000

(c) Number of data

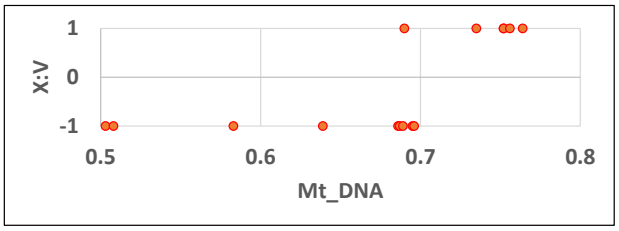
Metrics ＼ word order feature	S:V	O:V	X:V	N:P	G:N	A:N	D:N	Nu:N	R:N
S/H_rate	147	170	145	173	127	177	159	166	173
Y_DNA	163	167	81	162	164	169	144	152	158
Mt_DNA	19	23	15	20	21	23	17	22	22
S_ratio	164	172	129	167	171	181	156	156	167



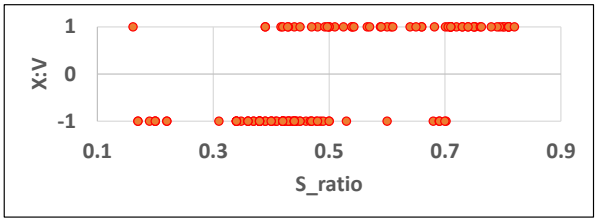
(a) S/H_rate



(b) Y_DNA



(c) Mt_DNA



(d) S_ratio

Figure 1 Scattering diagrams between X:V feature value (after quantification) and four thought pattern metrics

The cells colored pink are correlation coefficients that are significant at a significance level of 5%, and all of them are positive. This shows that head initial language speakers tend to score low on all four metrics and head final language speakers tend to score high on all four metrics.

[iii] All data used in this paper are open on the site: http://www.ne.jp/asahi/eharate/eharate/ronbun.files/NLP2024.xlsx.

[iv] https://sites.google.com/view/fukuimasayasu (accessed on 11 Feb. 2023)

Word order features which are significanct with all four thought pattern metrics are X:V and A:N. O:V has no significance in S/H_rate and Mt_DNA.

Figure 1 shows scattering diagrams of X:V feature values (after quantification) and four thought pattern metrics values. In all scattering diagrams, data points are distributed from the lower left to the upper right, reflecting the positive correlation coefficients.

4 Conclusion

Spearman's rank correlation coefficients are calculated between nine word order features and four thought pattern metrics. Results show correlation coefficients that are significant with a significance level of 5% are all positive. The word order features that are significant for all four thought pattern metrics are the word order of oblique and verb (X:V), and the word order of adjective and noun (A:N). The word order of object and verb (O:V), which we have focused on in previous studies, has no significance concerning the two metrics.

The relation between the four thought pattern metrics and the thought patterns themselves are mapped as shown in Table 4.

Table 4 Relation between four thought pattern metrics and the thought patterns themselves

Metrics	Thought pattern	
	Low	High
S/H_rate	Extrapunitive	Intropunitive
Y_DNA	Competitive	Cooperative
Mt_DNA	Competitive	Cooperative
S_ratio	Optimistic	Pessimistic

From Table 4, it can be seen that head initial language speakers tend to have extrapunitive, competitive and optimistic thought patterns, while head final language speakers tend to have intropunitive, cooperative and pessimistic thought patterns.

5 Related works and discussions

The relationship between language and thought is well known as the Sapir-Whorf hypothesis [11][12]. Although the relationship between the lexical structure of languages and thought patterns is extensively studied (for example [13][14]), there are few studies on the relationship between syntax, particularly word order, and thought patterns. A few examples, which are all qualitative studies, are the following:

Kido [15] claims that speakers who use narrow to wide word order in compound nouns expressing place names or person's names[v] tend to think with subject-oriented logic, and speakers who use the opposite word order tend to think with place-oriented logic.

Nishimitsu [16] claims that speakers of prepositional languages (PN order) tend to have an impulsive cognitive style and speakers of postpositional languages (NP order) tend to have a reflective cognitive style.

Shibata [17] claims that speakers who use lower digit to higher digit word order in compound numerals[vi] tend to have zoom-out type cognition and speakers who use opposite word order tend to have zoom-in type cognition.

In a bird language, it is said that important elements precede and unimportant elements follow [18]. If this is true in human languages, speakers of VX word order languages give more importance to their action (V) than to the surrounding environment (X). Therefore, they first decide their action (subject) and then think about the environment (place), so they have impulsive and zoom-out type of cognition. On the other hand, speakers of XV word order languages give more importance to the surrounding environment (X) than to their action (V). Therefore, they first think about the environment (place) and then decide their action (subject), so they have the reflective, zoom-in type of cognition.

[v] For the place name, the narrow place name precedes and the wide place name follows; for the person's name, the personal name precedes and the family name follows.

[vi] For example, in the case of compound numeral 14, 4 precedes and 10 follows (for example, "fourteen" in English).

References

1. Oka, Tomoyuki. Linguistics of Place. Tokyo: Hituzi Syobo Publishing. 2013. (in Japanese)
岡智之. 場所の言語学. 東京: ひつじ書房. 2013.
2. Li, Charles N. and Thompson, Sandra A. Subject and Topic: A New Typology of Language. In Charles N. Li. Subject and Topic. New York: Academic Press. 1976.
3. Dryer, Matthew S. and Haspelmath, Martin (eds.). The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology. 2013.
http://wals.info. (Accessed on 2018-03-09.).
4. de Marneffe, Marie-Catherine; Manning, Christopher; Nivre, Joakim; Zeman, Daniel. Universal Dependencies, Computational Linguistics, Vol. 47, No. 2, pp. 255-308. 2021.
5. Ehara, Terumasa. Relation between Word Order Parameters and Suicide / Homicide Rates. Jounal of Yamanashi Eiwa College, Vol.13, pp.9-29. 2015.
6. Ehara, Terumasa. Relation between Word Order of Languages and the Entropy of Y-chromosome Haplogroup Distribution of the Speakers' Population. Proceedings of the 24th Annual Meeting of the Association for Natural Language Processing, P10-8, pp.1019-1023. 2018.
7. Ehara, Terumasa. Relation between Word Order of Languages and the Entropy of Mitochondrial DNA Haplogroups Distribution of the Speakers' Population in Eurasia, Proceedings of The 26th Annual Meeting of The Association for Natural Language Processing, P5-7, pp.1197-1200. 2020.
8. Ehara, Terumasa. Relation between Word Order of Languages and Serotonin Transporter Gene, Proceedings of The 27th Annual Meeting of The Association for Natural Language Processing, P3-14, pp.579-584. 2021.
9. Yamamoto, Hideki. Survey and Historical Study of Geographical and Genealogical Distribution of Word Order around the World. Hiroshima: Keisuisha, 2003. (in Japanese).
山本秀樹. 世界諸言語の地理的・系統的語順分布とその変遷, 広島: 渓水社. 2003.
10. Fukui, Masayasu. College Analysis Manual. 2023. (in Japanese) (accessed on 2023-2-11)
福井正康. College Analysis 総合マニュアル. 2023. (2023 年 2 月 11 日アクセス)
11. Sapir, Edward. Language - An Introduction to the Study of Speech. New York: Harcourt, Brace & Co. Inc. 1956.
泉井久之助(訳). 言語. 東京: 紀伊国屋書店. 1957.
12. Whorf, Benjamin Lee. Language, Thought and Reality. Cambridge: The M.I.T. Press. 1956.
池上嘉彦(訳). 言語・思考・現実. 東京: 弘文堂. 1978.
13. Imai, Mutstumi. Language and Thought. Tokyo: Iwanami Shoten, Publishers. 2010. (in Japanese)
今井むつみ. ことばと思考. 東京: 岩波新書. 2010.
14. Guy Deutscher. Through the Language Glass: Why the World Looks Different in Other Languages. New York: Metropolitan Books, 2010.
椋田直子(訳). 言語が違えば、世界も違って見えるわけ. 東京: インターシフト. 2012.
15. Kido, Sessho. Philosophy of Place. Tokyo: Bungeisha. p.324 note 4-2-8. 2003. (in Japanese)
城戸雪照. 場所の哲学. 東京: 文芸社. p.324 note 4-2-8. 2003.
16. Nishimitsu, Yoshihiro. Cognitive Style and Language Typology. Kuroshio Linguistics School. 2012. (in Japanese)
西光義弘. 認知スタイルと言語類型, くろしお言語大学塾. 2012.
https://www.gengoj.com/seminar/view.php?seminar_list_id=8. (Accessed on 2012-5-6).
17. Shibata, Katsuyuki. History of Intracranial Conflict between Language and Cognition. IEICE Technical Report. TL2014-40. pp.13-18. 2014. (in Japanese)
柴田勝征. 言語 vs 認知の脳内抗争史. 信学技報. TL2014-40. pp.13-18. 2014.
18. Yamagiwa, Juichi and Suzuki, Toshitaka. What are the animals and birds talking about?. Tokyo: Shuueisha Publishing. pp.63-65. 2023. (in Japanese)
山際寿一, 鈴木俊貴. 動物たちは何をしゃべっているのか？. 東京: 集英社. pp.63-65. 2023.

Appendix Head final ratio r for X:V obtained from UDD

UDD_language_name	n(VX)	n(XV)	r
Assyrian	14	0	0.00
Beja	0	0	---
Madi	1	0	0.00
Makurap	4	0	0.00
Naija	0	0	---
Neapolitan	1	0	0.00
Tagalog	71	0	0.00
Coptic	4046	169	0.04
Manx	1127	41	0.04
Bambara	895	56	0.06
Ancient_Hebrew	2678	203	0.07
Cebuano	28	2	0.07
Irish	5642	467	0.08
Scottish_Gaelic	2712	223	0.08
Kiche	470	53	0.10
Arabic	10343	1230	0.11
English	29699	3869	0.12
Faroese	2892	376	0.12
Yoruba	383	51	0.12
Thai	1133	173	0.13
Javanese	104	19	0.15
South_Levantine_Arabic	23	4	0.15
Wolof	1535	276	0.15
French	27805	5137	0.16
Breton	443	88	0.17
Icelandic	71449	14133	0.17
Nheengatu	79	16	0.17
Norwegian	28571	5964	0.17
Yupik	55	11	0.17
Danish	4931	1116	0.18
Swedish	10046	2141	0.18
Indonesian	6604	1542	0.19
Hebrew	16411	4178	0.20
Swedish_Sign_Language	66	18	0.21
Welsh	789	212	0.21
Greek	2618	722	0.22
Ligurian	287	83	0.22
Portuguese	32396	8990	0.22
Spanish	36368	10318	0.22
Galician	5401	1613	0.23
Romanian	37929	11267	0.23
Albanian	38	12	0.24
Gheg	531	170	0.24
Karo	26	8	0.24
Old_Church_Slavonic	2757	880	0.24
Serbian	4403	1417	0.24
Guajajara	458	154	0.25
Italian	33187	10846	0.25
Vietnamese	569	189	0.25
Gothic	2734	952	0.26
Zaar	79	29	0.27
Catalan	14083	5812	0.29
North_Sami	1726	747	0.30
Polish	20248	8655	0.30
Western_Sierra_Puebla_Nahuatl	240	105	0.30
Maltese	1496	666	0.31
Croatian	7669	3555	0.32
Bulgarian	4211	2083	0.33
Old_East_Slavic	11696	5666	0.33
Ukrainian	5427	2624	0.33
Belarusian	11030	6103	0.36
Tupinamba	132	73	0.36
Russian	77211	44876	0.37
Finnish	11787	7319	0.38
Guarani	11	7	0.39
Kaapor	3	2	0.40
Pomak	1167	784	0.40
Slovak	3273	2161	0.40
Frisian_Dutch	72	50	0.41

UDD_language_name	n(VX)	n(XV)	r
Karelian	161	114	0.41
Livvi	79	56	0.41
Old_French	7218	5110	0.41
Slovenian	9439	6841	0.42
Erzya	691	523	0.43
Moksha	47	37	0.44
Ancient_Greek	12861	10416	0.45
Czech	59905	48347	0.45
Komi_Permyak	17	14	0.45
Mbya_Guarani	350	320	0.48
Apurina	9	9	0.50
Estonian	20662	21849	0.51
Skolt_Sami	14	15	0.52
Dutch	6774	7557	0.53
Komi_Zyrian	236	285	0.55
Latvian	7851	9877	0.56
Upper_Sorbian	225	291	0.56
Latin	22053	29089	0.57
Lithuanian	896	1221	0.58
Teko	30	41	0.58
Akuntsu	11	16	0.59
Classical_Chinese	1432	2134	0.60
Khunsari	2	3	0.60
Afrikaans	838	1340	0.62
Hindi_English	293	483	0.62
Munduruku	53	89	0.63
Low_Saxon	51	91	0.64
Western_Armenian	2712	5080	0.65
Armenian	2005	3811	0.66
Umbrian	31	62	0.67
Xavante	2	4	0.67
Korean	3252	6985	0.68
Cantonese	51	111	0.69
Buryat	16	37	0.70
German	65468	156418	0.70
Warlpiri	4	10	0.71
Basque	2185	5877	0.73
Hungarian	1013	2942	0.74
Chukchi	69	211	0.75
Swiss_German	13	52	0.80
Turkish_German	438	1706	0.80
Sanskrit	301	1596	0.84
Abaza	7	53	0.88
Chinese	635	5024	0.89
Akkadian	110	2457	0.96
Amharic	10	313	0.97
Hittite	3	91	0.97
Bhojpuri	3	185	0.98
Kangri	2	120	0.98
Marathi	4	157	0.98
Persian	606	34670	0.98
Tatar	3	195	0.98
Turkish	1058	43114	0.98
Telugu	2	295	0.99
Urdu	72	10023	0.99
Uyghur	20	3455	0.99
Bengali	0	7	1.00
Hindi	62	30440	1.00
Japanese	14	146107	1.00
Kazakh	0	711	1.00
Kurmanji	0	16	1.00
Malayalam	0	6	1.00
Nayini	0	5	1.00
Old_Turkish	0	8	1.00
Sinhala	0	13	1.00
Soi	0	4	1.00
Tamil	2	875	1.00
Xibe	2	569	1.00
Yakut	0	77	1.00