Word order of subject, object, oblique, and verb

主語、目的語、斜語と動詞の語順

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

Using the Universal Dependency Database (UDD), the word order of subject (S), object (O), oblique (X), and verb (V) is investigated. The order of three dependent elements (S, O, X) and head verb are examined in the database and the frequency of each word order patterns are counted. Although previous studies investigated the order of S, O, and V or O, X, and V, no studies have investigated the order of S, O, X, and V to the author's knowledge. The study results show the following: (1) Frequent word orders are SVOX, SXOV, SOVX, and SOXV. (2) Frequent verb positions are 2nd, 4th, 3rd, and 1st. (3) In the case of verb position is 2nd, 3rd, or 4th, the most frequent element of the 1st position is S. (4) In the case of verb position is 2nd or 3rd, the most frequent element of 4th position is X. (5) In the order of S, O, and X ignoring verb position, SOX, SXO, and XSO are most frequent. (6) O is often adjacent to V.

1 Introduction

The author has quantitatively investigated the relationship between the word order of languages and speakers' thought patterns [1]. Here, "word order (feature) " means the order of dependent and head in the dependency grammar [2]. For example, in the case of the word order feature of object (O) and verb (V), O is the dependent and V is the head. The notation O:V is used in the paper to express this word order feature. The feature values for O:V are OV or VO. In the former, V follows O (head final type), and in the latter, V precedes O (head initial type). The paper describes the feature values for S:O:X:V which is the order of three dependents: subject (S), object (O), and oblique (X) to the same head verb (V).

Although S:O:V and O:X:V features were examined in the World Atlas of Language Structures (WALS) [3], no studies have investigated S:O:X:V to the author's knowledge. Universal Dependency Database (UDD) [4][5] is used in the paper to count the frequency of the feature values of S:O:X:V. Although [6] made a word order study using UDD, it did not mention S:O:X:V.

2 Data

UDD (v2.14) [5] is used as the original data[i]. UDD includes the dependency tree bank data for 161 languages. Frequencies of values of the following word order features are counted: S:O:X:V, S:O:V, S:X:V, and O:X:V. In the case of multiple files for the same language, the total frequency of these files is counted. The reason why S:O:V, S:X:V, and O:X:V are examined is to compare the WALS data.

In the UDD, S:V, O:V, and X:V are described as nsubj, obj, and obl, respectively. We restrict the data having the part of speech of the dependent of nsubj, obj, and obl as a noun (NOUN), pronoun (PRON), and proper noun (PROPN) and also restrict the data having the part of speech of the head of nsubj, obj and, obl as a verb (VERB).

---

[i] Downloaded on September 19th, 2024 from the site listed in [5].

3 Frequency of the feature values for S:O:X:V

The (relative) frequencies and total number of data (Total) for S:O:X:V feature values for each language are shown in Table 10 and in the Appendix.

For example, Abkhaz has 5 data and relative frequencies are 0 for SOXV, 0.4 for SXOV, and so on. In the cases of 6 languages of which the total number of data is 0, relative frequencies are expressed by "----".

For each language, the most frequent feature value is determined. For example, Abkhaz has SXOV which has the most frequent value of 0.4. In the case of the same frequency for several feature values, the leftmost feature value in Table 10 or in the Appendix is adopted as the most frequent one. For example, Abaza has SXOV, OXSV, and OXVS which have the most frequent value of 0.286, so Abaza has SXOV as the most frequent feature value. Table 1 shows the distribution of the number of languages for each most frequent feature value. Table 1 also shows the number of languages that have feature values where the verb position is 4th (***V), 3rd (**V*), 2nd (*V**), or 1st (V***) which are the total number of each column. Table 1 also shows the number of languages that have the same order of S:O:X ignoring verb position which is the total number of each row.

Table 1 Distribution of the number of languages for the most frequent feature values

S:O:X:V	Freq	S:O:X:V	Freq	S:O:X:V	Freq	S:O:X:V	Freq	S:O:X
SOXV	12	SOVX	14	SVOX	63	VSOX	8	97
SXOV	28	SXVO	8	SVXO	4	VSXO	0	40
OSXV	0	OSVX	2	OVSX	2	VOSX	0	4
OXSV	0	OXVS	0	OVXS	0	VOXS	1	1
XSOV	3	XSVO	6	XVSO	0	VXSO	0	9
XOSV	4	XOVS	0	XVOS	0	VXOS	0	4
***V	47	**V*	30	*V**	69	V***	9	155

From Table 1, we can see the following:

(1) Frequent feature values are SVOX (63), SXOV (28), SOVX (14), and SOXV (12).

(2) Frequent verb positions are *V** (69), ***V (47), **V* (30), and V*** (9).

(3) Frequent 1st elements for ***V word order are S (40), X (7), and O (0).

(4) Frequent 1st elements for **V* word order are S (22), X (6), and O (2).

(5) Frequent 1st elements for *V** word order are S (67), X (0) and O (2).

(6) Frequent 4th elements for **V* word order are X (16), O (14), and S (0).

(7) Frequent 4th elements for *V** word order are X (65), O (4), and S (0).

(8) Frequent feature values for S:O:X are SOX (97), SXO (40), and XSO (9).

(9) O is often adjacent to V: adjacent (125) and non-adjacent (30).

4 Frequency of the feature values for S:O:V, S:X:V and O:X:V

Distribution of feature values for S:O:V, S:X:V and O:X:V are, also, examined by the UDD. The method of examination is the same as the case for S:O:X:V. The distributions of the number of languages for the most frequent feature values are listed in Table 2, Table 3, and Table 4. The distributions from WALS data are also listed in Table 2 and Table 4.

Table 2 Distribution of the number of languages for most frequent feature values for S:O:V

S:O:V	Freq(UDD)	Freq(WALS)
SOV	56	564
OSV	3	4
SVO	86	488
OVS	3	11
VSO	9	95
VOS	1	25

Table 3 Distribution of the number of languages for most frequent feature values for S:X:V

S:X:V	Freq(UDD)
SXV	44
XSV	10
SVX	89
XVS	1
VSX	13
VXS	0

Table 4 Distribution of the number of languages for most frequent feature values for O:X:V

O:X:V	Freq(UDD)	Freq(WALS)
OXV	11	27
XOV	40	48
OVX	16	45
XVO	20	3
VOX	70	210
VXO	3	0

From Tables 2, 3, and 4, we can see the following:

(1) Frequent feature values for S:O:V are SVO (86), SOV (56), and VSO (9) by UDD and SOV (564), SVO (488), and VSO (95) by WALS.

The reason why the rankings of SVO and SOV are

switched between UDD and WALS is that UDD consists of many Indo-European languages that have much SVO order. Actually, 45% of UDD languages are Indo-European and only 5% of WALS languages of which S:O:V feature values are described are Indo-European.

(2) Frequent feature values for S:X:V are SVX (89), SXV (44), and VSX (13) by UDD.

(3) Frequent feature values for O:X:V are VOX (70), XOV (40), XVO (20), and OVX (16) by UDD and VOX (210), XOV (48) and OVX (45) by WALS.

The number of XVO type languages is 20 by UDD (Table 5) and 3 by WALS (Cantonese, Hakka, and Mandarin). This point will be discussed in the next section.

Table 5 XVO languages in UDD

Cantonese, Chinese, Chukchi, Classical_Chinese, Czech, Erzya, Estonian, Georgian, Hungarian, Karelian, Komi_Zyrian, Latvian, Lithuanian, Livvi, Mbya_Guarani, Moksha, Slovak, Slovenian, Western_Sierra_Puebla_Nahuatl

5 Language-by-language comparison of S:O:V and O:X:V feature values with WALS

Language-by-language comparison results for S:O:V and O:X:V feature values between UDD and WALS are shown in Table 6.

In Table 6, "Disagree (NDO)" means WALS's "No Dominant Order (NDO)" feature value that disagrees with UDD's values. "No data" means no data in WALS or the total number of data is zero in UDD.

Table 6 Comparison results

	S:O:V	O:X:V
Agree	73	30
Disagree (NDO)	10	11
Disagree	6	5
No data	72	115
Total	161	161

Table 7 shows 6 languages that disagree with WALS for S:O:V.

Table 7 Disagree feature values for S:O:V

UDD name	WALS name	S:O:V(UDD)	S:O:V(WALS)	Total
Abkhaz	Abkhaz	OVS	SOV	44
Amharic	Amharic	OVS	SOV	503
Beja	Beja	SVO	SOV	58
Georgian	Georgian	SVO	SOV	78
Munduruku	Mundurukú	OVS	SOV	30
Tagalog	Tagalog	VOS	VSO	70

In Table 7, the total number of data is less than 100

except for Amharic. They have low reliability. In the Amharic case, the subject or object is sometimes expressed by the inflection or suffix of the verb. For example, the verb "አስበላሁት" consists of the verb stem "አስበላ", the morpheme expressing the subject "ሁ" and the expletive morpheme (expl) "ት". In the UDD's data for the Amharic, these morphemes are considered with the independent elements [7]. So the above example is considered with VS word order and OVS is the most frequent S:O:V feature value of the Amharic in UDD and disagrees with WALS's SOV.

Five languages that disagree with WALS for O:X:V feature values are listed in Table 8.

Table 8 Disagree feature values for O:X:V

UDD name	WALS name	O:X:V(UDD)	O:X:V(WALS)	Total
Estonian	Estonian	XVO	VOX	14026
Guarani	Guaraní	OVX	VOX	2
Komi_Zyrian	Komi-Zyrian	XVO	VOX	132
Mbya_Guarani	Guaraní	XVO	VOX	162
Yupik	Yupik (Siberian)	VOX	OVX	14

Three languages of which the total number of data is greater than 100 have feature values XVO in UDD and VOX in WALS. As described above, WALS has only 3 languages that have XVO order. However, 20 XVO type languages are found in this study. Table 9 shows relative frequencies for O:X:V feature values for Estonian, Komi_Zyrian, and Mbya_Guarani. Table 9, also, shows the English case. Although VOX has 0.79 which is most frequent in English, XVO has 0.15 which is 2nd ranked.

Table 9 Relative frequencies of O:X:V feature values for several languages

Language	Total	OXV	XOV	OVX	XVO	VOX	VXO
English	9416	0	0	0.039	0.15	0.79	0.019
Estonian	14026	0.114	0.111	0.138	0.214	0.212	0.211
Komi_Zyrian	132	0.098	0.152	0.197	0.273	0.167	0.114
Mbya_Guarani	162	0.08	0.309	0.086	0.407	0.105	0.012

6 Consideration for No Dominant Order

Although the most frequent feature value is considered in this research, it may be No Dominant Order (NDO) in the case of the frequency ratio of the 1st ranked and the 2nd ranked is greater than some threshold level. Actually, in the researches [6] [8], NDO is adopted for the frequency ratio is greater than 0.5. The relation between the agreement rate for UDD and WALS and the threshold
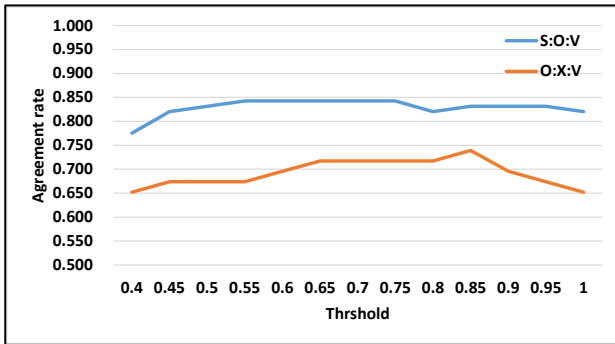
is examined. The result is shown in Figure 1.

Figure 1 Relation of agreement rate and the threshold

For S:O:V case, the agreement rate is highest (0.843) in the threshold range of 0.55 to 0.75. For O:X:V case, the agreement rate is highest (0.739) for the threshold of 0.85.

In the case of the threshold is 1, the agreement rates are the same as the rates calculated from the data in Table 6. They are 0.820 for S:O:V and 0.652 for O:X:V.

7 Conclusion

Using the Universal Dependency Database (UDD), the word order of Subject (S), Object (O), Oblique (X), and Verb (V) for 161 languages are examined. The result shows most frequent orders are SVOX, SXOV, SOVX, and SOXV. The word order of S, O, and V and O, X, and V are, also, examined and the results are compared with the World Atlas of Language Structure's data.

These results are descriptive and the explanative studies have remained as the future works.

Table 10 Distribution of the relative frequencies of S:O:X:V feature values with total number of data and maximum feature value (continue to the Appendix)

Lang name	Total	SOXV	SXOV	OSXV	OXSV	XSOV	XOSV	SOVX	SXVO	OSVX	OXVS	XSVO	XOVS	SVOX	SVXO	OVSX	OVXS	XVSO	XVOS	VSOX	VSXO	VOSX	VOXS	VXSO	VXOS	MAX
Abaza	7	0.143	0.286	0	0.286	0	0	0	0	0	0.286	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Abkhaz	5	0	0.4	0	0	0	0	0	0.2	0	0	0.2	0.2	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Afrikaans	599	0.295	0.182	0.068	0	0.105	0	0.255	0	0.013	0	0	0	0.042	0.007	0	0	0.02	0	0.012	0	0	0	0	0	SOXV
Akkadian	269	0.353	0.271	0.056	0.078	0.033	0.045	0.033	0.063	0.022	0	0.041	0	0.004	0	0	0	0	0	0	0	0	0	0	0	SOXV
Akuntsu	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	OVSX
Albanian	7	0	0	0	0	0	0	0	0	0	0	0.286	0	0.429	0.286	0	0	0	0	0	0	0	0	0	0	SVOX
Amharic	78	0.128	0.269	0	0.013	0	0.051	0.013	0.026	0.013	0.128	0.026	0.218	0.013	0	0.013	0	0.064	0.013	0.013	0	0	0	0	0	SXOV
Ancient_Greek	2821	0.05	0.079	0.036	0.02	0.044	0.016	0.103	0.095	0.07	0.029	0.048	0.023	0.125	0.062	0.043	0.018	0.019	0.019	0.043	0.013	0.029	0.004	0.009	0.001	SVOX
Ancient_Hebrew	225	0	0	0	0	0	0	0	0	0.009	0	0	0	0.133	0.053	0.013	0.004	0.036	0.036	0.431	0.116	0.111	0.004	0.053	0	VSOX
Apurina	10	0	0	0	0	0	0	0	0.2	0	0	0.1	0	0.6	0.1	0	0	0	0	0	0	0	0	0	0	SVOX
Arabic	1944	0	0	0	0	0	0	0	0.004	0	0	0.018	0	0.246	0.068	0.006	0.001	0.069	0.007	0.262	0.121	0.158	0.023	0.015	0.001	VSOX
Armenian	629	0.025	0.113	0.008	0.006	0.099	0.008	0.221	0.202	0.033	0.006	0.153	0.038	0.067	0.005	0.013	0.003	0	0	0	0	0	0	0	0	SOVX
Assyrian	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	SVOX
Azerbaijani	11	0.364	0.545	0	0	0.091	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Bambara	181	0	0	0	0	0.105	0	0.895	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Basque	915	0.051	0.066	0	0.001	0.104	0.001	0.326	0.098	0.012	0.021	0.035	0.086	0.068	0.022	0.057	0.015	0.02	0.009	0.003	0.001	0.003	0.001	0	0	SOVX
Bavarian	136	0.199	0.154	0.081	0	0.096	0.007	0.007	0	0.007	0	0	0	0.066	0.066	0.037	0	0.147	0	0.059	0.074	0	0	0	0	SOXV
Beja	3	0	0	0	0	0	0.333	0	0.333	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	XOSV
Belarusian	2077	0.002	0.002	0.002	0.003	0.006	0.003	0.016	0.107	0.062	0.015	0.247	0.018	0.357	0.104	0.015	0.011	0.001	0.019	0.002	0	0.001	0.004	0	0	SVOX
Bengali	2	0	0.5	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Bhojpuri	22	0.182	0.636	0	0	0.182	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Bororo	44	0	0	0	0	0.045	0	0.864	0	0.091	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Breton	40	0	0	0	0	0	0	0	0.025	0	0.025	0	0	0.35	0.05	0.075	0	0.1	0	0.3	0.05	0.025	0	0	0	SVOX
Bulgarian	1019	0	0.002	0	0	0.008	0	0.052	0.045	0.015	0.002	0.305	0.019	0.419	0.07	0.044	0.011	0	0.006	0	0	0.002	0.001	0	0	SVOX
Buryat	3	0.333	0	0	0	0.333	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Cantonese	46	0	0	0	0.022	0	0	0	0.5	0	0	0.348	0	0.087	0.043	0	0	0	0	0	0	0	0	0	0	SXVO
Cappadocian	18	0	0	0	0	0	0	0.056	0.056	0	0	0.389	0	0.167	0.056	0.167	0	0	0.056	0.056	0	0	0	0	0	XSVO
Catalan	9186	0.003	0.008	0.001	0	0.033	0	0.275	0.026	0.022	0.011	0.099	0.104	0.236	0.102	0.055	0.012	0.002	0.004	0.001	0.001	0.002	0.002	0	0.001	SOVX
Cebuano	13	0	0	0	0	0	0	0.077	0	0	0	0	0.077	0	0	0	0	0.077	0	0.385	0	0.385	0	0	0	VSOX
Chinese	1846	0.001	0.003	0.001	0	0.003	0	0.001	0.645	0.001	0	0.321	0	0.016	0.009	0	0	0	0	0	0.001	0	0	0	0	SXVO
Chukchi	6	0.333	0	0	0	0	0	0.167	0.167	0	0.167	0.167	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Classical_Armenian	272	0.015	0.033	0.007	0	0.004	0	0.081	0.07	0.029	0	0.015	0.007	0.426	0.04	0.081	0.033	0.051	0.011	0.033	0.007	0.037	0	0.015	0.004	SVOX
Classical_Chinese	2001	0.001	0.001	0	0	0.005	0	0.004	0.418	0	0	0.354	0	0.215	0.001	0	0	0	0	0	0	0	0	0	0	SXVO
Coptic	1005	0	0	0	0	0	0	0	0.001	0.001	0	0.064	0	0.831	0.103	0	0	0	0	0	0	0	0	0	0	SVOX
Croatian	1651	0.012	0.004	0.008	0.004	0.01	0.005	0.061	0.183	0.043	0.018	0.173	0.028	0.35	0.064	0.016	0.001	0.001	0.003	0.002	0.004	0.001	0	0.001	0	SVOX
Czech	14519	0.016	0.015	0.008	0.006	0.01	0.008	0.047	0.128	0.028	0.029	0.082	0.023	0.148	0.175	0.052	0.04	0.09	0.034	0.017	0.024	0.003	0.006	0.006	0.004	SVXO
Danish	1183	0.001	0.001	0.003	0	0	0	0.003	0.109	0.055	0.001	0.074	0.001	0.546	0.031	0.014	0	0.089	0	0.057	0.016	0	0	0	0	SVOX
Dutch	4530	0.113	0.135	0.039	0	0.059	0	0.11	0	0.013	0	0	0	0.173	0.069	0.021	0	0.147	0	0.081	0.036	0	0.002	0	0	SVOX
Egyptian	66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.758	0.136	0.091	0	0.015	0	VSOX
English	5860	0	0	0	0.001	0	0.001	0	0.018	0.057	0	0.199	0	0.707	0.016	0.001	0	0	0	0	0	0	0	0	0	SVOX
Erzya	127	0.024	0.024	0.008	0	0.024	0	0.126	0.094	0.016	0.016	0.126	0	0.252	0.205	0.024	0.008	0	0.024	0.008	0.024	0	0	0	0	SVOX
Estonian	5942	0.044	0.06	0.029	0	0.023	0.001	0.014	0.033	0.013	0.006	0.031	0.009	0.22	0.208	0.021	0.016	0.11	0.035	0.043	0.069	0.003	0.005	0.002	0.008	SVOX
Faroese	496	0	0	0.002	0	0	0.002	0.008	0.028	0.05	0.002	0.04	0.018	0.526	0.014	0.06	0	0.062	0	0.171	0.006	0	0	0	0	SVOX
Finnish	2498	0.002	0.002	0.011	0	0.003	0	0.004	0.028	0.03	0.002	0.201	0.013	0.446	0.23	0.008	0.01	0.001	0.002	0.004	0.001	0.001	0	0	0	SVOX
French	6049	0.002	0.001	0.001	0	0.021	0	0.135	0.032	0.046	0	0.209	0.001	0.435	0.103	0.007	0.001	0.001	0.002	0.003	0	0	0	0	0	SVOX
Frisian_Dutch	29	0.069	0.207	0.034	0	0.138	0.034	0.138	0	0.034	0	0	0	0.138	0	0.034	0	0.103	0	0.034	0.034	0	0	0	0	SXOV
Galician	1734	0.002	0.009	0.001	0.002	0.018	0.001	0.176	0.047	0.021	0.002	0.189	0.018	0.31	0.111	0.019	0.012	0.02	0.008	0.009	0.005	0.006	0.007	0.005	0.002	SVOX
Georgian	39	0.026	0.051	0.026	0	0.103	0.026	0.026	0.077	0.026	0	0.154	0	0.154	0.077	0.154	0	0.051	0.026	0	0	0	0	0	0.026	XSVO
German	76088	0.179	0.19	0.06	0.006	0.116	0.009	0.008	0	0.001	0	0	0	0.085	0.081	0.031	0.004	0.111	0.015	0.041	0.043	0.011	0.003	0.004	0	SXOV
Gheg	77	0.013	0.039	0	0	0.091	0	0.156	0.117	0.013	0	0.169	0.039	0.299	0.026	0.026	0	0.013	0	0	0	0	0	0	0	SVOX
Gothic	328	0.021	0.046	0.009	0.003	0.015	0	0.131	0.061	0.052	0.003	0.021	0.006	0.302	0.107	0.043	0.03	0.012	0.015	0.03	0.018	0.049	0.003	0.018	0.003	SVOX
Greek	454	0	0.004	0	0	0.002	0	0.088	0.055	0.009	0	0.198	0.015	0.37	0.117	0.075	0.02	0.007	0.011	0.015	0.007	0.004	0.002	0	0	SVOX
Guajajara	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Guarani	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Gujarati	14	0.286	0.5	0	0	0.143	0	0	0	0	0	0.071	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Haitian_Creole	45	0	0	0	0	0	0	0	0.022	0.133	0	0.111	0	0.644	0.089	0	0	0	0	0	0	0	0	0	0	SVOX
Hausa	12	0	0	0	0	0	0	0	0	0	0	0.5	0	0.417	0.083	0	0	0	0	0	0	0	0	0	0	XSVO
Hebrew	1980	0	0	0	0	0	0	0.011	0.006	0.001	0	0.098	0.001	0.345	0.174	0.022	0.005	0.195	0.01	0.084	0.033	0.006	0.001	0.01	0.001	SVOX
Highland_Puebla_Nahuatl	24	0	0	0	0	0	0	0	0.042	0.333	0	0.292	0	0.208	0.125	0	0	0	0	0	0	0	0	0	0	OSVX
Hindi	4922	0.113	0.466	0.027	0.013	0.328	0.018	0.001	0.014	0	0	0.018	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV

References

[1] EHARA, Terumasa. Oblique and verb word order relates to speakers' thought patterns, **Proceedings of the 30th Annual Meeting of The Association for Natural Language Processing**, P1-27, pp.285-290, 2024.

[2] Kodama, Tokumi. Izon bunpou no kenkyuu. Tokyo: **Kenkyusha Shuppan**, 1987 (in Japanese).

[3] Dryer, Matthew S.; Haspelmath, Martin (Eds.). The World Atlas of Language Structures Online. Leipzig: **Max Planck Institute for Evolutionary Anthropology**. 2013.

[4] de Marneffe, Marie-Catherine; Manning, Christopher; Nivre, Joakim; Zeman, Daniel. Universal Dependencies, **Computational Linguistics**, Vol. 47, No. 2, pp. 255-308. 2021

[5] Zeman, Daniel; et al. Universal Dependencies 2.15, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), **Faculty of Mathematics and Physics, Charles University**, 2024.
http://hdl.handle.net/11234/1-5787.

[6] Choi, Hee-Soo; Guillaume, Bruno; Fort, Karën. Corpus-based Language Universals Analysis using Universal Dependencies, **SyntaxFest Quasy 2021 - Quantitative Syntax**, Mar 2022, Sofia, Bulgaria. ffhal-03501774v2ff, 2022.

[7] Seyouml, Binyam Ephrem; Miyao, Yusuke; Mekonnen, Baye Yimam. Universal Dependencies for Amharic, **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, pp. 2216–2222, Miyazaki, Japan: European Language Resources Association (ELRA), 2018.

[8] Dryer, Matthew S. Determining dominant word order. In Matthew S. Dryer and Martin Haspelmath, (Eds), The World Atlas of Language Structures Online. Leipzig: **Max Planck Institute for Evolutionary Anthropology**, 2013.
https://wals.info/chapter/s6

Lang name	Total	SOXV	SXOV	OSXV	OXSV	XSOV	XOSV	SOVX	SXVO	OSVX	OXVS	XSVO	XOVS	SVOX	SVXO	OVSX	OVXS	XVSO	XVOS	VSOX	VSXO	VOSX	VOXS	VXSO	VXOS	MAX
Hittite	11	0.545	0.273	0.091	0.091	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Hungarian	669	0.018	0.088	0.027	0.013	0.057	0.007	0.126	0.241	0.016	0.018	0.115	0.039	0.096	0.039	0.013	0.012	0.022	0.028	0	0	0.006	0.012	0.003	0.003	SXVO
Icelandic	13099	0.012	0.004	0.007	0	0.007	0.001	0.023	0.033	0.035	0.001	0.028	0.001	0.459	0.07	0.034	0.001	0.07	0.002	0.174	0.031	0.003	0.001	0.001	0	SVOX
Indonesian	1451	0	0	0	0	0	0.002	0	0.057	0.019	0	0.38	0	0.531	0.012	0	0	0	0	0	0	0.001	0	0	0	SVOX
Irish	751	0	0	0	0	0	0	0	0.001	0	0	0.017	0	0.003	0	0	0	0.108	0.001	0.754	0.116	0	0	0	0	VSOX
Italian	6474	0.006	0.015	0.005	0.002	0.014	0.001	0.072	0.08	0.036	0.002	0.221	0.007	0.352	0.141	0.015	0.005	0.004	0.012	0.001	0	0.004	0.003	0	0.001	SVOX
Japanese	8374	0.154	0.539	0.015	0.005	0.276	0.011	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Javanese	82	0	0	0	0	0	0.012	0	0.024	0.024	0	0.39	0	0.549	0	0	0	0	0	0	0	0	0	0	0	SVOX
Kaapor	3	0.333	0	0	0.333	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Kangri	16	0.125	0.312	0	0	0.5	0.062	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XSOV
Karelian	54	0	0	0.019	0	0	0	0	0.019	0.019	0	0.315	0.019	0.333	0.241	0	0	0	0.037	0	0	0	0	0	0	SVOX
Karo	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Kazakh	88	0.148	0.489	0.011	0.011	0.307	0.034	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Khunsari	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XOSV
Kiche	29	0	0	0	0	0	0	0	0	0	0	0.103	0	0.552	0.034	0	0	0	0.034	0	0.172	0.103	0	0	0	SVOX
Komi_Permyak	5	0	0.2	0	0	0	0.2	0	0	0	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0	0	SVXO
Komi_Zyrian	55	0.055	0.018	0	0.055	0	0	0.182	0.164	0	0	0.164	0.018	0.145	0.091	0.018	0.036	0	0	0.036	0	0.018	0	0	0	SOVX
Korean	573	0.113	0.379	0.003	0.003	0.264	0.003	0.229	0	0.005	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Kurmanji	16	0.062	0.062	0	0	0	0	0.875	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Kyrgyz	98	0.173	0.531	0	0	0.224	0.061	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Latgalian	3	0	0.333	0	0	0.333	0	0	0	0	0	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	SOXV
Latin	6301	0.094	0.112	0.055	0.02	0.056	0.06	0.108	0.129	0.033	0.008	0.03	0.009	0.128	0.067	0.022	0.007	0.011	0.006	0.016	0.009	0.012	0.002	0.005	0.002	SXVO
Latvian	1798	0.012	0.017	0.012	0.006	0.028	0.006	0.102	0.22	0.055	0.046	0.208	0.069	0.166	0.043	0.007	0.002	0.001	0.002	0	0	0.001	0	0	0	SXVO
Ligurian	46	0	0.017	0	0	0.152	0	0.304	0.022	0.022	0	0.13	0	0.304	0.022	0.022	0	0	0	0	0	0.022	0	0	0	SOVX
Lithuanian	189	0.021	0.011	0.016	0	0.021	0.005	0.111	0.19	0.042	0.085	0.19	0.053	0.143	0.079	0.011	0.011	0	0.005	0	0	0	0	0	0	SXVO
Livvi	14	0	0.071	0	0	0	0	0.071	0.214	0	0.071	0.071	0	0.357	0.071	0	0	0	0.071	0	0	0	0	0	0	SVOX
Low_Saxon	293	0.249	0.137	0.096	0	0.024	0.007	0.058	0.007	0.01	0	0	0	0.157	0.038	0.034	0	0.068	0	0.082	0.024	0.007	0	0.003	0	SOXV
Luxembourgish	3	0.667	0	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Macedonian	3	0	0	0	0	0	0	0.333	0	0	0	0	0	0.667	0	0	0	0	0	0	0	0	0	0	0	SVOX
Madi	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	OSVX
Maghrebi_Arabic_French	109	0	0	0	0	0	0	0.046	0.009	0	0	0.009	0	0.734	0.183	0.018	0	0	0	0	0	0	0	0	0	SVOX
Makurap	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Malayalam	25	0.08	0.2	0.08	0	0.44	0.16	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XSOV
Maltese	157	0	0	0	0	0	0	0	0.096	0.013	0	0.344	0.006	0.414	0.121	0	0	0.006	0	0	0	0	0	0	0	SVOX
Manx	397	0	0	0	0	0	0	0	0	0	0	0	0	0.005	0	0.005	0	0	0	0.841	0.149	0	0	0	0	VSOX
Marathi	21	0.381	0.381	0.048	0	0.19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Mbya_Guarani	64	0.062	0.016	0	0.016	0.219	0.016	0.031	0.016	0.031	0	0.453	0	0.109	0.016	0.016	0	0	0	0	0	0	0	0	0	XSVO
Middle_French	134	0	0.007	0	0	0.03	0	0.112	0.037	0.224	0	0.164	0	0.246	0.104	0.015	0	0	0	0.037	0.015	0	0.007	0	0	SVOX
Moksha	27	0	0	0	0.037	0.037	0.037	0.037	0.148	0	0	0	0	0.222	0.37	0	0.037	0.037	0.037	0	0	0	0	0	0	SVXO
Munduruku	18	0	0	0	0	0	0	0.111	0.056	0	0	0	0	0	0	0.778	0.056	0	0	0	0	0	0	0	0	OVSX
Naija	855	0	0	0	0	0	0	0	0.006	0.001	0	0.194	0	0.773	0.026	0	0	0	0	0	0	0	0	0	0	SVOX
Nayini	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XOSV
Neapolitan	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	SVOX
Nheengatu	95	0	0	0	0	0	0	0.011	0.032	0.042	0	0.179	0	0.547	0.168	0	0	0	0.021	0	0	0	0	0	0	SVOX
North_Sami	509	0.002	0.02	0.006	0	0.002	0	0.01	0.065	0.051	0.004	0.088	0.002	0.554	0.145	0	0.004	0.018	0.002	0.012	0.01	0.002	0.002	0	0.002	SVOX
Norwegian	6066	0	0	0	0	0	0	0.066	0.026	0	0.07	0	0	0.566	0.044	0.022	0	0.123	0.041	0.068	0.012	0	0	0	0	SVOX
Old_Church_Slavonic	972	0.05	0.066	0.02	0.012	0.028	0.004	0.088	0.107	0.033	0.012	0.041	0.01	0.221	0.072	0.039	0.032	0.019	0.034	0.03	0.009	0.048	0.004	0.01	0.009	SVOX
Old_East_Slavic	3833	0.051	0.085	0.017	0.01	0.045	0.011	0.091	0.058	0.028	0.013	0.051	0.008	0.151	0.078	0.032	0.006	0.06	0.038	0.058	0.032	0.039	0.015	0.02	0.003	SVOX
Old_French	2333	0.024	0.065	0.017	0.001	0.061	0.013	0.249	0.042	0.076	0.003	0.064	0.033	0.165	0.063	0.044	0.009	0.03	0.004	0.024	0.005	0.005	0.001	0.001	0.001	SOVX
Old_Irish	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Old_Turkish	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Ottoman_Turkish	39	0.179	0.513	0	0	0.282	0.026	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Paumari	0	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----
Persian	7186	0.402	0.434	0.013	0.002	0.134	0.004	0.009	0.001	0	0.001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Polish	2751	0.005	0.005	0.003	0.001	0.005	0	0.019	0.086	0.017	0.01	0.187	0.024	0.395	0.154	0.01	0.015	0.008	0.022	0.01	0.005	0.008	0.007	0	0.003	SVOX
Pomak	118	0	0.025	0	0.008	0.034	0.008	0.144	0.068	0.025	0.017	0.068	0.008	0.178	0.102	0.153	0.076	0.008	0	0.034	0.008	0.025	0.008	0	0	SVOX
Portuguese	8195	0	0.003	0.001	0.001	0.008	0.001	0.035	0.042	0.031	0	0.243	0.004	0.498	0.104	0.005	0.001	0.009	0.006	0.004	0.001	0.002	0	0	0	SVOX
Romanian	5050	0.005	0.032	0.001	0	0.024	0.002	0.13	0.043	0.011	0.002	0.109	0.029	0.306	0.076	0.096	0.022	0.014	0.009	0.044	0.012	0.022	0.007	0.004	0.001	SVOX
Russian	11338	0.008	0.007	0.008	0.004	0.011	0.005	0.024	0.109	0.058	0.013	0.262	0.026	0.28	0.123	0.012	0.013	0.005	0.016	0.005	0.002	0.002	0.002	0.001	0.001	SVOX
Sanskrit	1036	0.167	0.204	0.092	0.04	0.12	0.036	0.104	0.022	0.042	0.014	0.02	0.017	0.021	0.023	0.018	0.008	0.006	0.01	0.009	0.005	0.009	0.007	0.003	0.005	SXOV
Scottish_Gaelic	574	0	0	0	0	0	0	0	0	0	0	0	0	0.002	0	0.007	0	0.044	0	0.699	0.249	0	0	0	0	VSOX
Serbian	811	0.001	0	0.002	0	0.002	0	0.02	0.134	0.009	0.004	0.195	0.006	0.533	0.08	0.012	0	0	0	0	0	0	0	0	0	SVOX
Sinhala	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Skolt_Sami	9	0.111	0	0	0	0	0	0.222	0.111	0	0	0	0	0.444	0.111	0	0	0	0	0	0	0	0	0	0	SVOX
Slovak	672	0.018	0.018	0.001	0.007	0.009	0.015	0.107	0.15	0.033	0.028	0.113	0.052	0.164	0.144	0.022	0.013	0.058	0.012	0.006	0.015	0.003	0.003	0.001	0.006	SVOX
Slovenian	1806	0.014	0.006	0.02	0.003	0.011	0.011	0.102	0.231	0.075	0.075	0.138	0.053	0.117	0.045	0.034	0.013	0.017	0.011	0.007	0.01	0.003	0.001	0.001	0.002	SXVO
Soi	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XOSV
South_Levantine_Arabic	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	SVOX
Spanish	7895	0	0.005	0.003	0.001	0.008	0	0.06	0.075	0.031	0.009	0.194	0.016	0.401	0.148	0.028	0.007	0.003	0.007	0.001	0.001	0.002	0.002	0.002	0.001	SVOX
Swedish	2922	0	0.001	0	0	0	0	0	0.069	0.051	0	0.064	0	0.589	0.026	0.017	0	0.09	0.005	0.068	0.01	0.007	0	0	0	SVOX
Swedish_Sign_Language	3	0	0	0	0	0	0	0	0.333	0	0	0	0	0.667	0	0	0	0	0	0	0	0	0	0	0	SVOX
Swiss_German	14	0.286	0.214	0.071	0	0.286	0	0	0	0	0	0	0	0.143	0	0	0	0	0	0	0	0	0	0	0	SOXV
Tagalog	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.294	0.059	0.294	0.353	0	0	VOXS
Tamil	101	0.05	0.347	0.089	0.02	0.366	0.119	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	XSOV
Tatar	18	0.167	0.389	0.056	0	0.389	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Teko	4	0	0	0	0	0.25	0	0	0	0	0	0	0	0.5	0.25	0	0	0	0	0	0	0	0	0	0	SVOX
Telugu	46	0.239	0.478	0.087	0.065	0.065	0	0.043	0.022	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Telugu_English	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Thai	163	0	0	0	0	0	0	0.006	0	0	0	0.35	0	0.632	0.012	0	0	0	0	0	0	0	0	0	0	SVOX
Tswana	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	SVOX
Tupinamba	3	0	0	0	0	0	0	0.333	0	0	0	0	0	0	0	0	0	0	0	0.333	0.333	0	0	0	0	SOVX
Turkish	3267	0.239	0.53	0.013	0.012	0.155	0.012	0.01	0.009	0.001	0.002	0.002	0.006	0.002	0.001	0.001	0.001	0.001	0	0.001	0.001	0	0.001	0	0	SXOV
Turkish_German	286	0.122	0.217	0.066	0.003	0.185	0.003	0.049	0.014	0.014	0	0.038	0.024	0.066	0.035	0	0.059	0.003	0.024	0	0	0	0	0	0	SXOV
Ukrainian	915	0.012	0.007	0.007	0.004	0.013	0.002	0.026	0.103	0.058	0.009	0.243	0.036	0.287	0.149	0.009	0.009	0.001	0.014	0.002	0.001	0.004	0.001	0	0.003	SVOX
Umbrian	2	0.5	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOXV
Upper_Sorbian	71	0.127	0.028	0.042	0	0.07	0	0	0.028	0	0	0.014	0	0.169	0.197	0.056	0.014	0.197	0	0.028	0.028	0	0	0	0	SVXO
Urdu	1543	0.136	0.506	0.023	0.01	0.283	0.019	0.005	0.008	0.001	0	0.008	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Uyghur	298	0.319	0.456	0.013	0.007	0.198	0.007	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Veps	13	0	0	0	0	0	0	0.077	0.077	0	0.154	0	0	0.615	0.077	0	0	0	0	0	0	0	0	0	0	SVOX
Vietnamese	498	0	0	0.002	0	0	0	0.002	0.056	0.006	0	0.536	0	0.357	0.04	0	0	0	0	0	0	0	0	0	0	XSVO
Warlpiri	2	0	0	0	0	0	0	0.5	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Welsh	45	0	0	0	0	0	0	0	0	0	0	0.022	0	0	0	0.067	0	0.089	0	0.689	0	0.022	0	0	0	VSOX
Western_Armenian	920	0.034	0.143	0.007	0.003	0.098	0.001	0.239	0.203	0.014	0.005	0.133	0.027	0.063	0.01	0.011	0.004	0	0.003	0	0	0	0	0.001	0	SOVX
Western_Sierra_Puebla_Nahuatl	10	0	0	0	0	0	0	0.1	0	0	0	0.5	0.3	0	0	0	0	0	0	0.1	0	0	0	0	0	XSVO
Wolof	408	0	0.002	0	0	0.037	0.017	0.14	0.017	0.181	0	0.164	0	0.397	0.044	0	0	0	0	0	0	0	0	0	0	SVOX
Xavante	3	0	0.333	0	0	0	0	0.667	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SOVX
Xibe	73	0.11	0.548	0.027	0	0.301	0	0.014	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Yakut	10	0.2	0.6	0.1	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	SXOV
Yoruba	130	0.008	0	0.015	0	0	0	0.008	0.038	0.054	0	0.085	0	0.723	0.069	0	0	0	0	0	0	0	0	0	0	SVOX
Yupik	5	0	0	0	0	0	0	0.2	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0	0	SVOX
Zaar	11	0	0	0	0	0	0	0	0	0	0	0.091	0	0.909	0	0	0	0	0	0	0	0	0	0	0	SVOX