

Word order characteristics analyzed by Hayashi's quantification method type III 数量化理論 III 類を用いた語順類型の分析

Terumasa EHARA
江原暉将

Yamanashi Eiwa College
山梨英和大学

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

1 Introduction

There are many languages in the world. These languages have similarity and difference each other. Linguistic typology intends to clarify these characteristics. The word order characteristics are one of the main topics in this field [Greenburg, 1966], [Comrie, 1981], [Hawkins, 1983], [Tsunoda, 1991], [Matsumoto, 2006].

This paper presents word order characteristics analyzed by Hayashi's quantification method type III (HQM). Our previous works [Ehara, 1995; Ehara, 2007] used multi-dimensional scaling (MDS) to quantify the characteristics. In the previous works, word order characteristics were quantified as follows: +10: head final order, -10: head initial order and 0: no dominant order. This quantification is "arbitrary". On the other hand, HQM makes non-arbitrary quantification.

2 Word order characteristics

As [Ehara, 2007], we select 7 characteristics¹ from 13 word order characteristics used in the WALS [Dryer, 2005]. They are listed in Table 1. We define three values for each feature: head initial (Initial), head final (Final) and no dominant order (Both). For example, in the case of "Order of subject (S) and verb (V)"; "VS(head initial)", "SV(head final)" and "No dominant order" are defined. The languages which have another feature values are not used in our

¹ We use "characteristics" and "features" as the same meaning.

analysis. In other words, we only consider languages which have one of the three feature values for all 7 features. We have obtained 576 languages from [Dryer, 2005] by this filtering. The number of languages which have Initial (I), Both² (B) and Final (F) feature values are, also, listed in Table 1.

Table 1 Word order characteristics

No.	Word order characteristics	Final (F)	Initial (I)	# of I	# of B	# of F
1	Subject(S) and Verb(V) in a declarative sentence	VS	SV	84	33	459
2	Object(O) and Verb(V) in a declarative sentence	VO	OV	307	28	241
3	Noun(N) and Adposition(Ap)	Ap-N	N-Ap	275	30	271
4	Genitive(G) and Noun(N)	NG	GN	227	38	311
5	Adjective(A) and Noun(N)	NA	AN	377	30	169
6	Determiner(D) and Noun(N)	N-Dm	Dm-N	305	0	271
7	Numeral(Nm) and Noun(N)	N-Nm	Nm-N	281	27	268

3 Analysis by the Hayashi's quantification method type III

To make experiments by Hayashi's quantification method type III, we use Hayakari's free tool [Hayakari, 2007].

As the result, accumulated contribution ratio is shown in Figure 1.

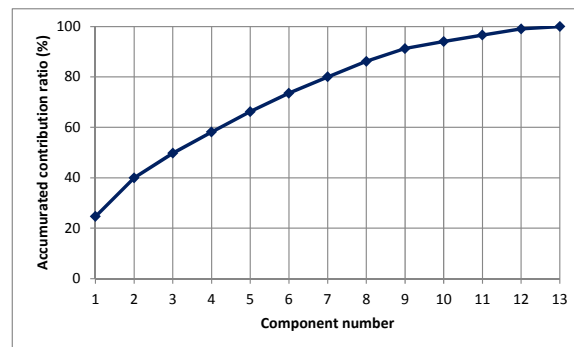


Figure 1 Accumulated contribution ratio

² "Both" means "no dominant order".

Up to the second component, accumulated contribution ratio reaches 40%. Our previous work by MDS results 75% accumulated contribution ratio up to the second component.

Table 2 and Figure 2 show first and second components of all feature values for seven features in the Table 1.

Table 2 First and second components of all feature values

Feature number	Feature value	First component	Second component
1	initial	-1.432	2.171
	both	0.008	2.753
	final	0.261	-0.595
2	initial	-1.169	0.452
	both	0.506	0.816
	final	1.429	-0.670
3	initial	-1.305	0.548
	both	-0.219	-0.474
	final	1.349	-0.504
4	initial	-1.457	0.384
	both	-0.492	1.704
	final	1.124	-0.489
5	initial	-0.578	-0.779
	both	0.253	0.857
	final	1.244	1.585
6	initial	-0.954	-0.938
	final	1.074	1.055
7	initial	-0.357	-1.585
	both	-0.336	0.435
	final	0.408	1.618

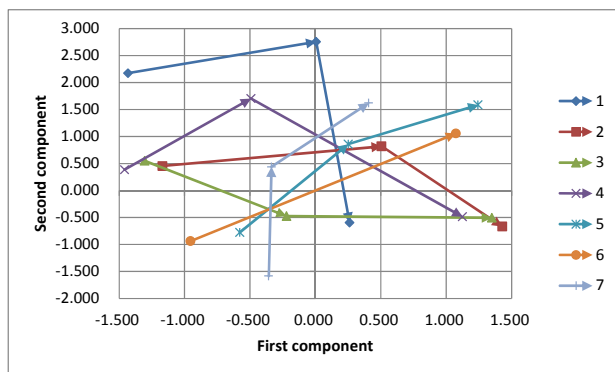


Figure 2 First and second components of all feature values (arrows are from Initial to Both and Both to Final)

We can see that first components of initial feature values are all negative and these of final feature values are all positive. The first components of both feature values are between the first components of initial feature values and these of final feature

values. For the second components, initial feature value and final feature value have opposite sign. For features 1 to 4, initial feature values are positive and final feature values are negative. However, for features 5 to 7, initial feature values are negative and final feature values are positive.

The first and second components of 576 languages are shown in Figure 3 with OV, VO and Both feature values of the feature number 2.

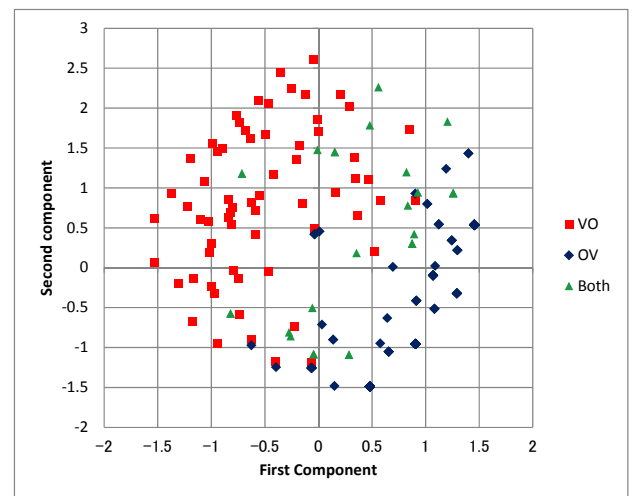


Figure 3 First and second components of all languages with VO(Initial), OV(Final) and Both feature values of the feature number 2

4 Comparison between HQM results and MDS results

We examine components of HQM and MDS for languages. The relation between them is shown in Figure 4 and Figure 5. Correlation coefficient between the first components of the two methods is 0.996 and correlation coefficient between the second components of the two methods is -0.973. They are highly correlated.

5 Relation between HQM results and suicide rate and homicide rate

We have made researches about the relation between word order features and suicide rate and homicide rate [Ehara, 2013]. Two word order features are used in this analysis. They are order of object and noun

(feature number 2) and order of adjective and noun (feature number 5). Feature values of “initial” and “final” for above two features are used. “No dominant order” value is not considered in this analysis. 64 languages are analyzed in the previous work.

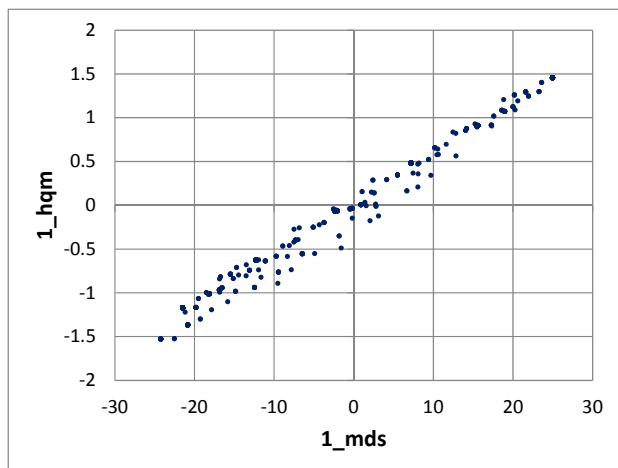


Figure 4 Relation between the first components of HQM and MDS

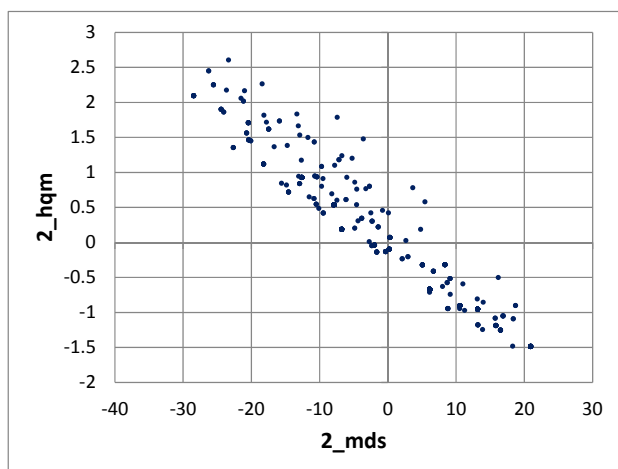


Figure 5 Relation between the second components of HQM and MDS

Table 3 First and second components of restricted feature and feature values

Feature number	Feature value	First component	Second component
2	initial	-0.984	-0.984
	final	1.0163	1.0163
5	initial	-0.66	0.66
	final	1.5152	-1.5152

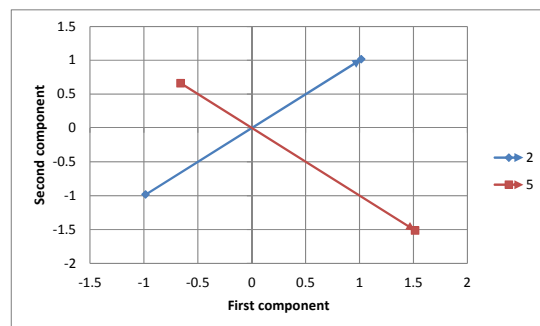


Figure 6 First and second components of restricted feature and feature values (arrows are from Initial to Final)

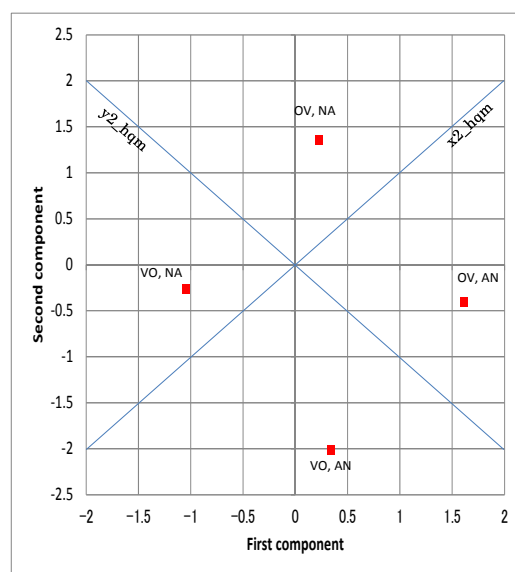


Figure 7 First and second components of 64 languages

Here, we make HQM to these data. Results are shown in Table 3, Figure 6 and Figure 7.

From Figure 7, we can see VO-OV direction and NA-AN direction are almost orthogonal and slanted with 45 degrees from the original axes. So, we rotate axes 45 degrees to $x2_hqm$ and $y2_hqm$.

Then multiple regression analysis is done comparing [Ehara, 2013]. Criterion variables are suicide rate ($\log_{10}(S)$) and homicide rate ($\log_{10}(H)$). Explanatory variables are GDP per capita (GDP), average annual temperature (TMP), average annual precipitation (PRC), $x2_hqm$ and $y2_hqm$. Results are shown in Table 4 and Table 5. Contribution ratio for suicide rate is 0.2779

and contribution ratio for homicide rate is 0.3153. They are almost equal to the results obtained by the previous study.

Table 4 Result of multiple regression analysis for suicide rate

log10(S)	Partial regression coefficient	Standardized partial regression coefficient	T value	Degree of freedom	Porbability	Correlation coefficient	Partial correlation coefficient
GDP	-0.045	-0.079	-0.513	59.000	0.610	0.213	-0.067
TMP	-0.012	-0.229	-1.271	59.000	0.209	-0.268	-0.163
PRC	0.002	0.385	2.865	59.000	0.006	0.122	0.350
x2_hqm	-0.076	-0.188	-1.412	59.000	0.163	-0.139	-0.181
y2_hqm	-0.142	-0.416	-2.882	59.000	0.006	-0.386	-0.351
Intercept	0.990	0.000	2.579	59.000	0.012		

Table 5 Result of multiple regression analysis for homicide rate

log10(H)	Partial regression coefficient	Standardized partial regression coefficient	T value	Degree of freedom	Porbability	Correlation coefficient	Partial correlation coefficient
GDP	-0.449	-0.631	-4.185	59.000	0.000	-0.499	-0.479
TMP	-0.001	-0.020	-0.113	59.000	0.911	0.323	-0.015
PRC	0.001	0.203	1.550	59.000	0.126	0.226	0.198
x2_hqm	-0.119	-0.231	-1.786	59.000	0.079	0.097	-0.227
y2_hqm	-0.044	-0.101	-0.720	59.000	0.474	0.162	-0.093
Intercept	2.093	0.000	4.415	59.000	0.000		

From the t-test, PRC and y2_hqm have non-zero partial regression coefficient for suicide rate with less than 1% significance level. For homicide rate, GDP has non-zero partial regression coefficient with less than 1% significance level. X2_hqm has also non-zero partial regression coefficient but its significance level is almost 8%.

Partial regression coefficient of y2_hqm for suicide rate is negative. It means that suicide rate decreases from AN to NA. Partial regression coefficient of x2_hqm for homicide rate is also negative. It means that homicide rate tends to decrease from VO to OV. These results are similar to the previous study.

6 Conclusion

To analyze word order characteristics, Hayashi's quantification method type III (HQM) is used. Our previous study uses multi-dimensional scaling (MDS) which has arbitrariness to quantify the word order characteristics. HQM is non-arbitral quantification method. We make two experiments. The first experiment is quantification using 7 word order

characteristics. We examine first component and second component of HQM and MDS results. Both components are highly correlated each other. The second experiment is regression analysis for suicide rate and homicide rate. The results by HQM are also similar to the previous results.

References

- [Comrie, 1981] Comrie, Bernard: Language Universals and Linguistic Typology, University of Chicago Press, 1981.
- [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, 1995 (in Japanese).
江原暉将: 多次元尺度構成法を用いた語順パラメータの間の関係付け, 言語処理学会第1回年次大会発表論文集, pp.173-176, 1995.
- [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, pp.15-18, 2007.
- [Ehara, 2013] Ehara, Terumasa: Relation between the word order characteristics and suicide/homicide rates (4), Proceedings of The 19th Annual Meeting of The Association for Natural Language Processing, pp.826-829, 2013.
- [Greenburg, 1966] Greenburg, Joseph: Language universals, with special reference to feature hierarchies, Janua Linguarum, Series Minor 59, Mouton, 1966.
- [Hawkins, 1983] Hawkins, John A.: Word Order Universals, Academic Press, 1983.
- [Hayakari, 2007] Hayakari, Susumu: <http://www7b.biglobe.ne.jp/~hayakari/3Quantify.htm>, 2007.
- [Matsumoto, 2006] Matsumoto, Katsumi: A Worldwide Perspective on Languages - Historical Linguistics and Linguistic Typology -, Sanseido, 2006 (in Japanese).
松本克己: 世界言語への視座 - 歴史言語学と言語類型論 -, 三省堂, 2006.
- [Tsunoda, 1991] Tsunoda, Tasaku: Languages of the World and Japanese, Kurosio Shuppan, 1991 (in Japanese).
角田太作: 世界の言語と日本語, くろしお出版, 1991.