Examining MDD and MHD as Syntactic Complexity Measures with Intermediate Japanese Learner Corpus Data

Saeko Komori (Chubu University , Japan)

Masatoshi Sugiura (Nagoya University , Japan)

Wenping Li (Dalian Maritime University, China)

Table of Contents

1 Introduction

- This study examines :

two syntactic complexity measures, MDD and MHD

for Japanese language development

with NS and NNS written corpus data

2 Previous studies on syntactic complexity

Ortega (2015) overviewed recent SLA writing and syntactic complexity studies, and discussed some factors that might affect differences in results across studies:

1) A factor of measurement

- Subordination measures

- Length-based measures

- Frequency-based measures

2) Another factor of genre differences

These are some of the factors that might lead to inconclusive results across studies.

2 Previous studies on MDD and MHD

	Study	MDD/MHD	Language	NS/NNS
1	Jing and Liu (2015)	MDD and MHD	English and Czech	NS
2	Liu et al. (2017)	MDD	20 natural languages	NS
3	Ouyang and Jiang (2017)	MDD	English as a second language	NNS
4	Komori et al. (2018, 2019)	MDD and MHD	Japanese	NS/NNS

2.1 Jing and Liu (2015)

- Proposed two "statistical metrics" (MDD and MHD) to predict the structural complexity of language

- compared two SVO languages
  - English: rigid word order and
  - Czech: relatively free word order

Main findings:

- There are significantly positive correlations between SL, MDD, and MHD.

- For longer sentences,

  English prefers to increase the MDD, while

  Czech tends to enhance the MHD.

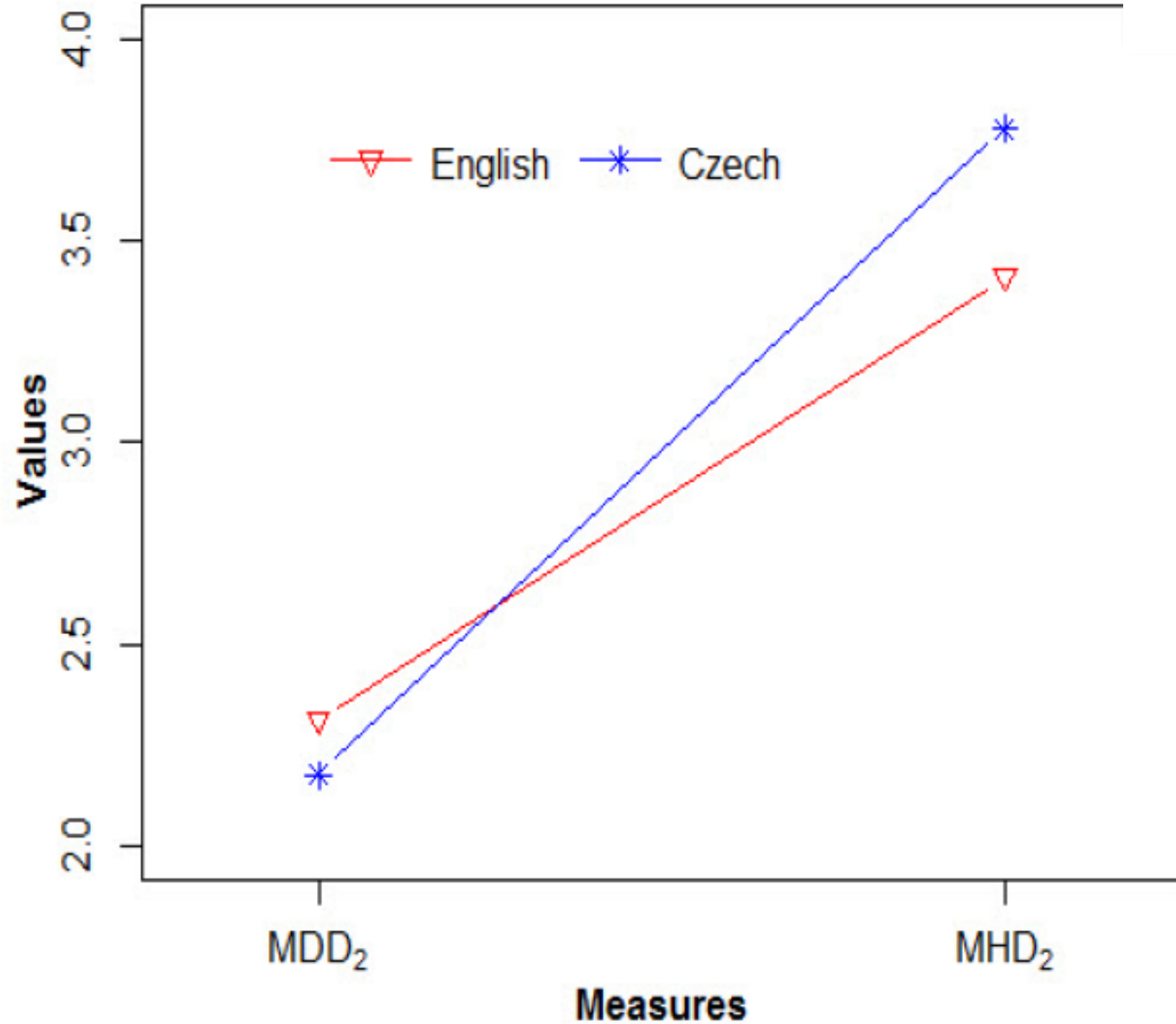| Lang	X-Y	Cor	p	k	R^2
	SL-MDD	0.54	<0.01	0.03	0.3
en	SL-MHD	0.74	<0.01	0.09	0.54
	MDD-MHD	0.19	<0.01	0.41	0.04
	SL-MDD	0.42	<0.01	0.02	0.18
cs	SL-MHD	0.74	<0.01	0.12	0.54
	MDD-MHD	0.11	<0.01	0.36	0.01

Table 3: Correlations between SL, MDD, and MHD

Table 3 and Figure 7 of Jing and Liu (2015)

2.2 Liu et al. (2017)

2.3 Jiang and Ouyang (2017)

MDD at Different Grades

Junior high school: The MDD of Chinese EFL learners' English writings **increases significantly** (p=0.000) from J1 (1.841) to J2 (2.061), but **stays stable** (p=0.936>0.05) from J2 (2.061) to J3 (2.064).

Senior high school: The MDD of Chinese EFL learner' English writings **first increases significantly** (p=0.003) at S1 (2.188), then **continues increasing insignificantly** (p=0.445>0.05) at S2, but experiences **a significant (p=0.022) decrease** at S3 (2.125).

University: the MDD of their writings **increases significantly** (p=0.000) at first, but then **keeps steady** (p=0.782>0.005).

Slide from Ouyang and Jiang (2018)

2.4 Advanced Japanese learners' study (Komori et al., 2018 and 2019)

YNUs	CL	CM	CH	NS
MDD	2.16	2.08	2.16	2.07
MHD	1.75	1.84	1.98	1.97
words	8806	10525	10810	9022
sentences	1316	1523	1391	1209
DD sum	16150	18715	20304	16160
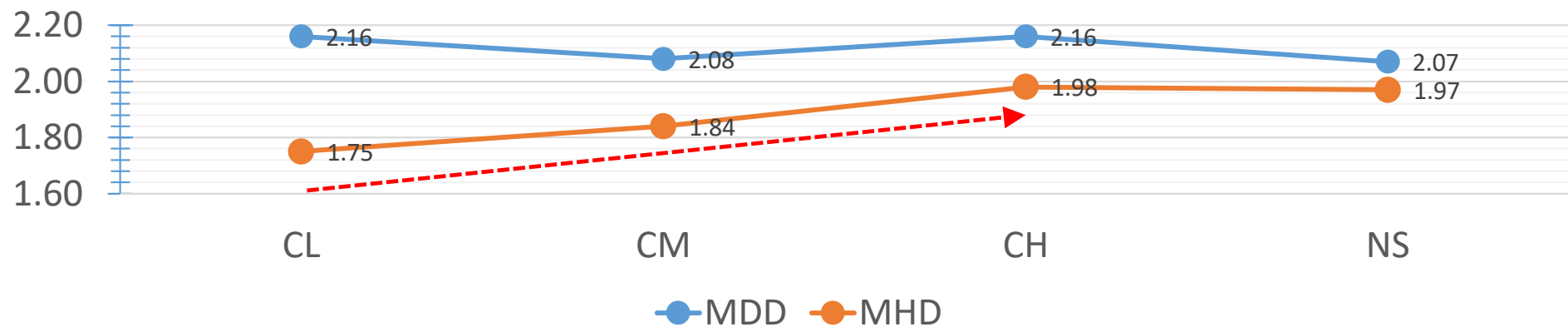
MHD: gradual increase as learning level rises

3 Research Question

"Can we use MDD and MHD in order to measure Japanese learners' syntactic complexity development using intermediate learners' corpus data?"

4 Analysis
4.1 Procedure of calculation of MDD and MHD:

1. parse each sentence syntactically with Cabocha, a Japanese dependency structure analyzer (Kudo and Matsumoto, 2002) and IPADic.

2. edit the data by retrieving dependent ID and governor ID.

3. use the dependent ID and governor ID to calculate the dependent distance (DD).

4. calculate MDD and MHD

Parsing

- Cabocha 0.69 + IPADic 2.7.0

Dependent word ID

Governor word ID

* 0 1D 0/0 1.748680↓
この　　　連体詞,*,*,*,*,*,この,コノ,コノ↓

* 1 6D 0/1 -1.501487↓
度　　　名詞,非自立,副詞可能,*,*,*,度,タビ,タビ↓
は　　　助詞,係助詞,*,*,*,*,は,ハ,ワ↓

* 2 3D 2/3 2.307352↓
お　　　接頭詞,名詞接続,*,*,*,*,お,オ,オ↓
伺い　　　名詞,一般,*,*,*,*,伺い,ウカガイ,ウカガイ↓
し　　　動詞,自立,*,*,サ変・スル,連用形,する,シ,シ↓
たい　　　助動詞,*,*,*,特殊・タイ,基本形,たい,タイ,タイ↓

* 3 4D 0/1 1.218672↓
こと　　　名詞,非自立,一般,*,*,*,こと,コト,コト↓
が　　　助詞,格助詞,一般,*,*,*,が,ガ,ガ↓

* 4 6D 0/0 -1.501487↓
あり　　　動詞,自立,*,*,五段・ラ行,連用形,ある,アリ,アリ↓

* 5 6D 0/1 -1.501487↓
メール　　名詞,サ変接続,*,*,*,*,メール,メール,メール↓
を　　　助詞,格助詞,一般,*,*,*,を,ヲ,ヲ↓

* 6 -1D 0/5 0.000000↓
送ら　　　動詞,自立,*,*,五段・ラ行,未然形,送る,オクラ,オクラ↓
せ　　　動詞,接尾,*,*,一段,連用形,せる,セ,セ↓
て　　　助詞,接続助詞,*,*,*,*,て,テ,テ↓
いただき　　　　動詞,非自立,*,*,五段・カ行イ音便,連用形,いただく,イタダキ,イタダキ↓
まし　　　助動詞,*,*,*,特殊・マス,連用形,ます,マシ,マシ↓
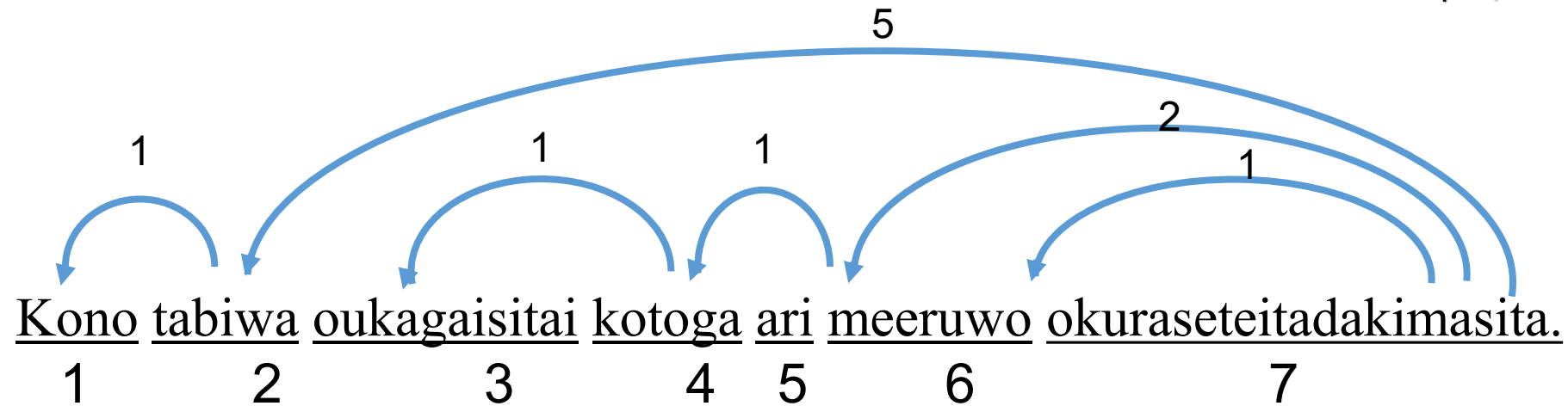た　　　助動詞,*,*,*,特殊・タ,基本形,た,タ,タ↓
。　　　記号,句点,*,*,*,*,。,。,。　↓

Example sentence:

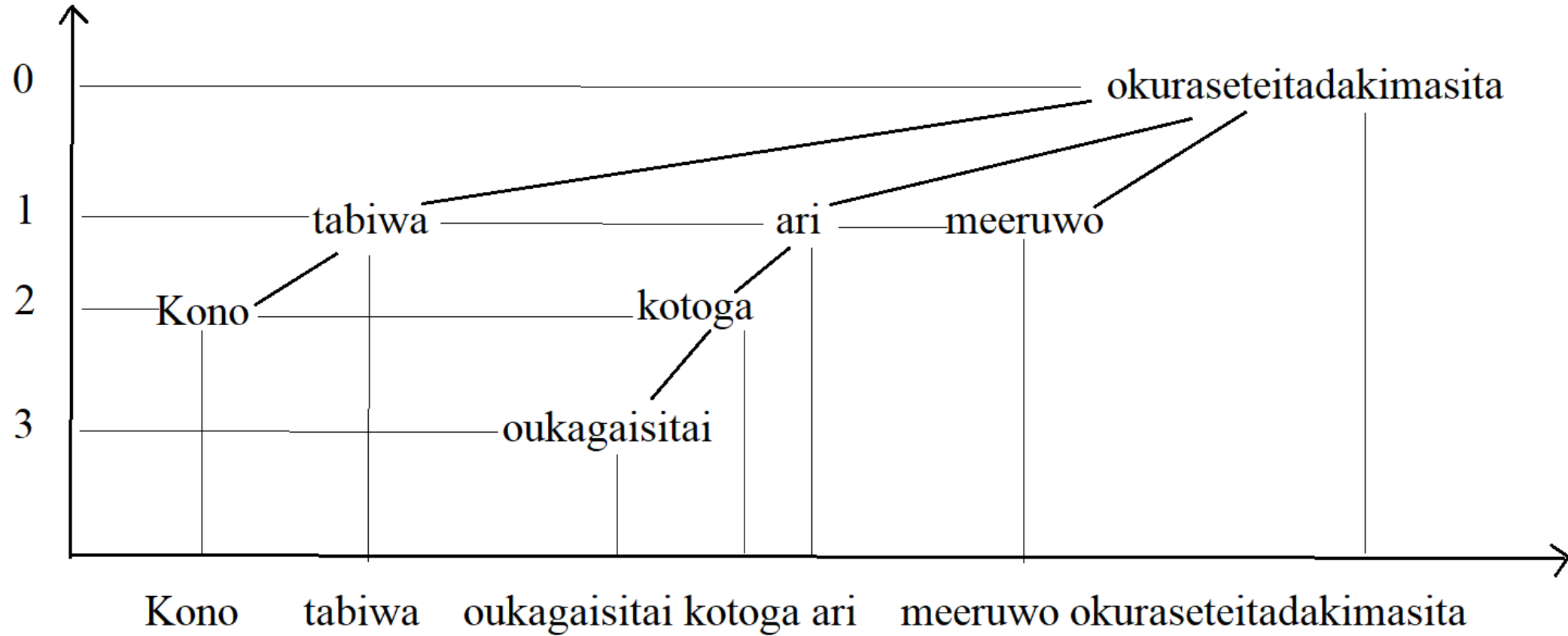"Kono tabiwa oukagaisitai kotoga ari, meeruwo okuraseteitadakimasita."

(I sent an email because I have something that I would like to ask you.)

$$\text{MDD} = (1+5+1+1+2+1) \div (7-1) = 1.83$$

Example of Dependency tree and MHD calculation

HD = 2 + 1 + 3 + 2 + 1 + 1
MHD = HD / (V - 1)
    = 10 / 6
    = 1.67

4.2 Data:
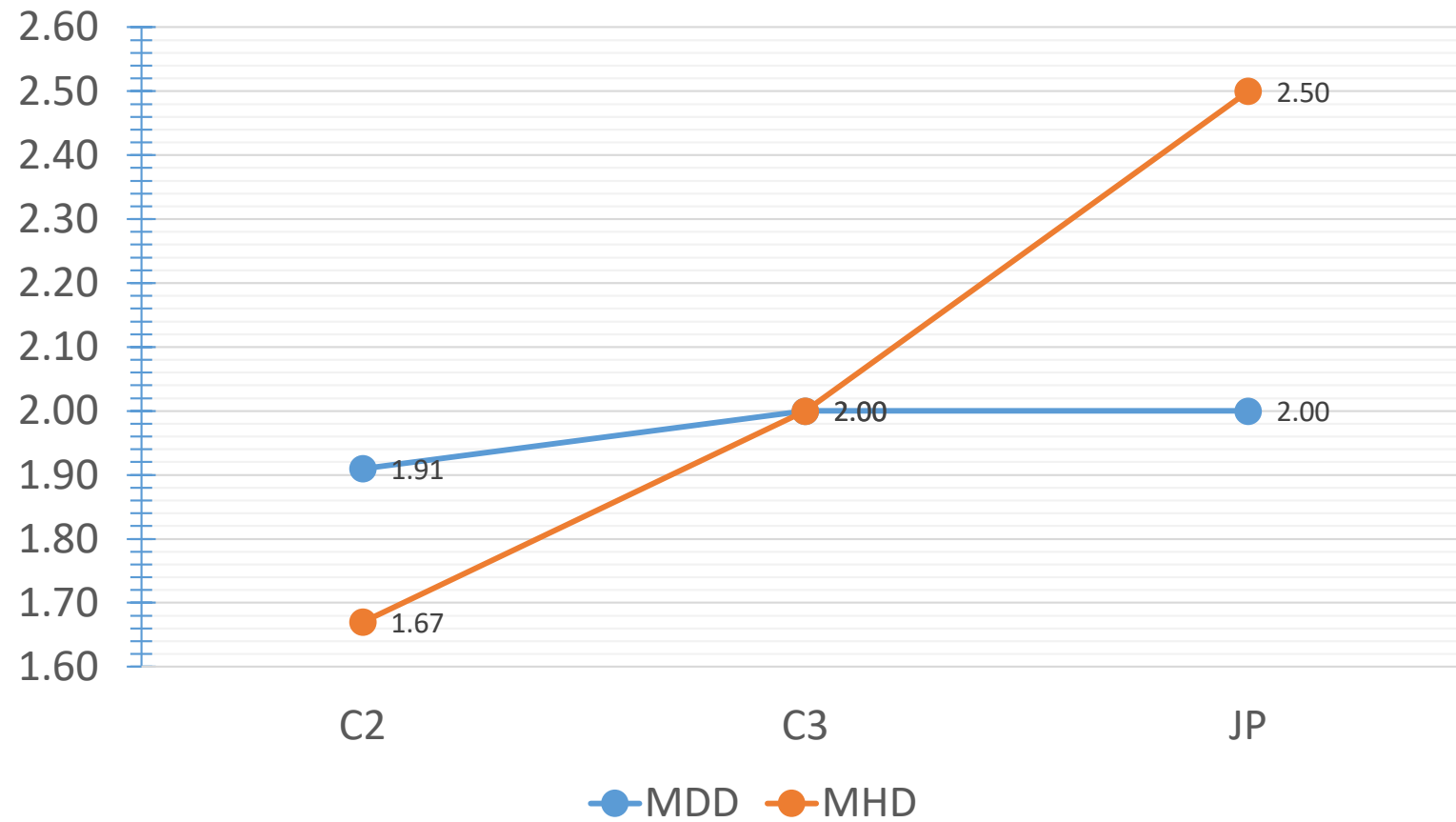# Intermediate learners and native speakers of Japanese

Group	Participants	Sentences	Words
C2 (second year university learners)	38	721	10,296
C3 (third year university learners)	33	605	11,786
JP (Japanese university students)	35	463	12,495

- Topic: "Will you decide your plans for life after graduation by yourself or will you consult other people?"
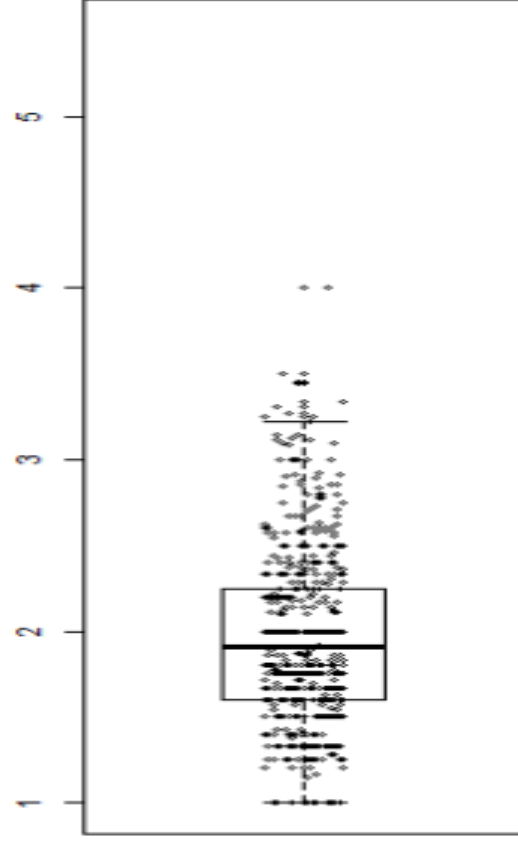
5 Results: SL, MDD and MHD

Group	Number of Sentences	Median		
		Sentence Length (SL) (Min, Max)	MDD (Min, Max)	MHD (Min, Max)
C2	592	6	1.91	1.67
		(4, 4)	(1.00, 4.00）	(1.00, 4.00)
C3	547	8	2.00	2.00
		(4, 18)	(1.00, 4.21)	(1.00, 4.64)
JP	429	10	2.00	2.50
		(4, 24)	(1.00, 3.96)	(1.00, 8.17)

MDD and MHD

MDD

MHD
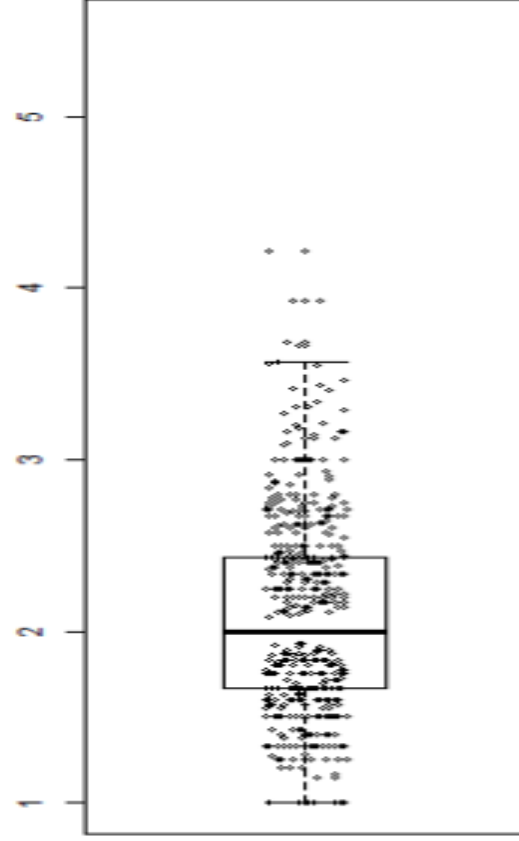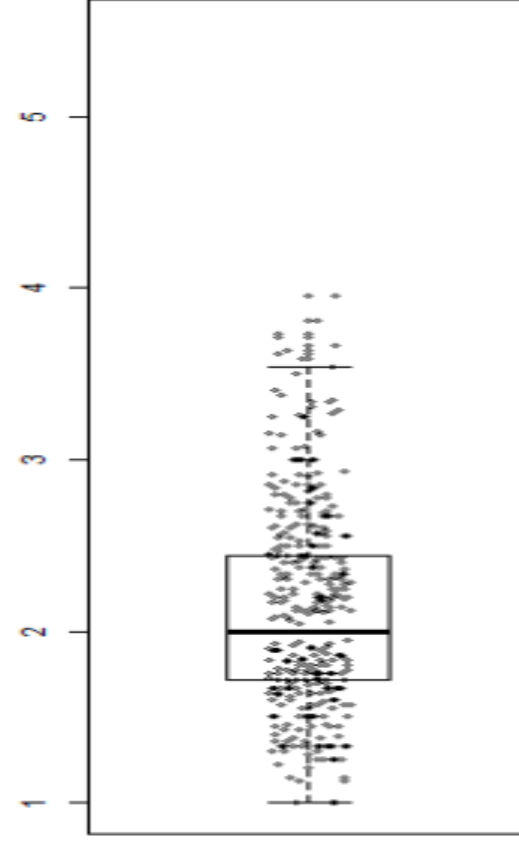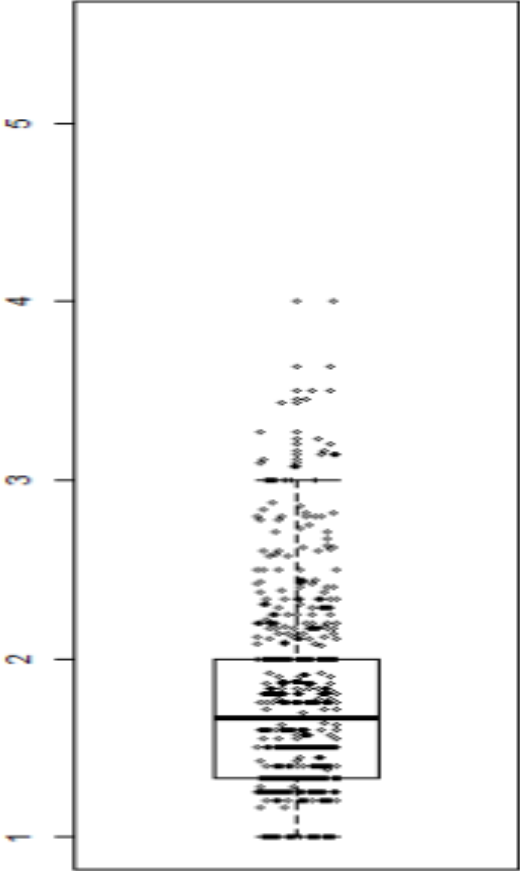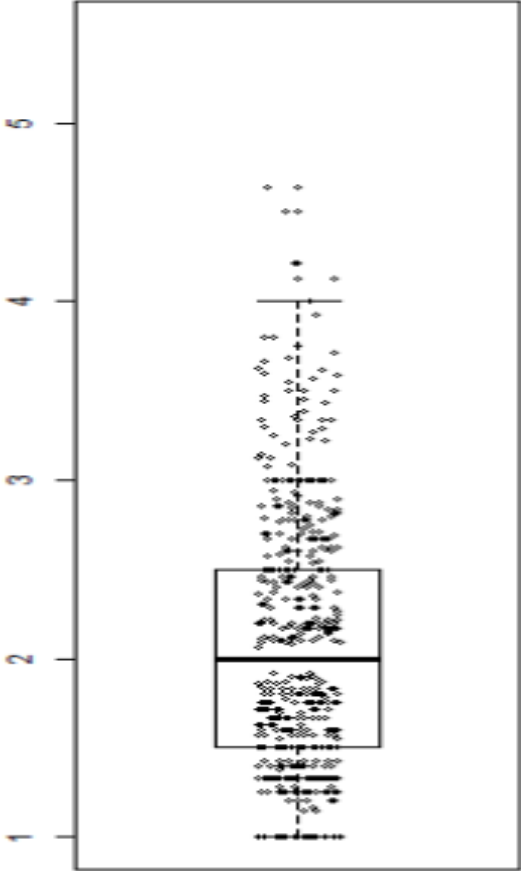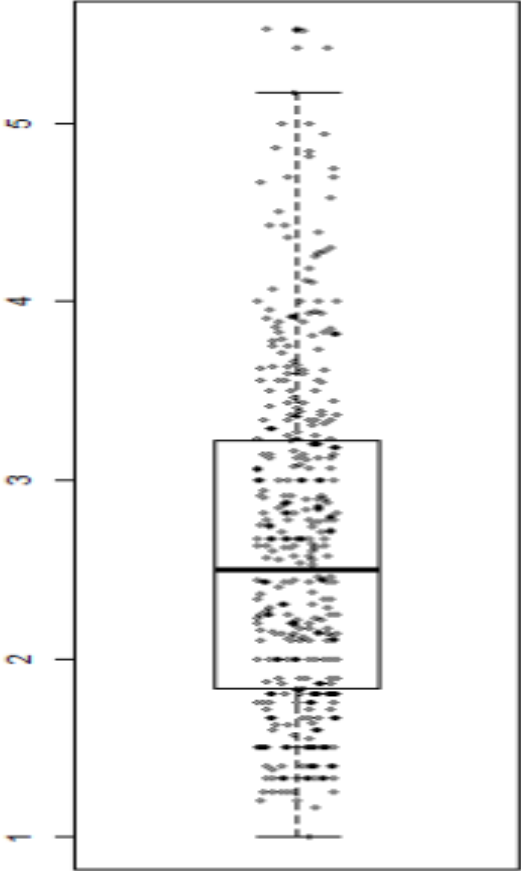
Brunner-Munzel Test and Cliff's delta of the MDD and MHD

	MDD		
	BM	*p*	Cliff's delta
C2 v. C3	3.88	.0001	.13 (negligible)
C3 v. JP	1.04	.2988	.04 (negligible)
C2 v. JP	4.86	<.0001	.17 (small)

	MHD		
	BM	*p*	Cliff's delta
C2 v. C3	7.73	<.0001	.25 (small)
C3 v. JP	10.26	<.0001	.35 (medium)
C2 v. JP	19.22	<.0001	.56 (large)

Correlations between SL, MDD and MHD
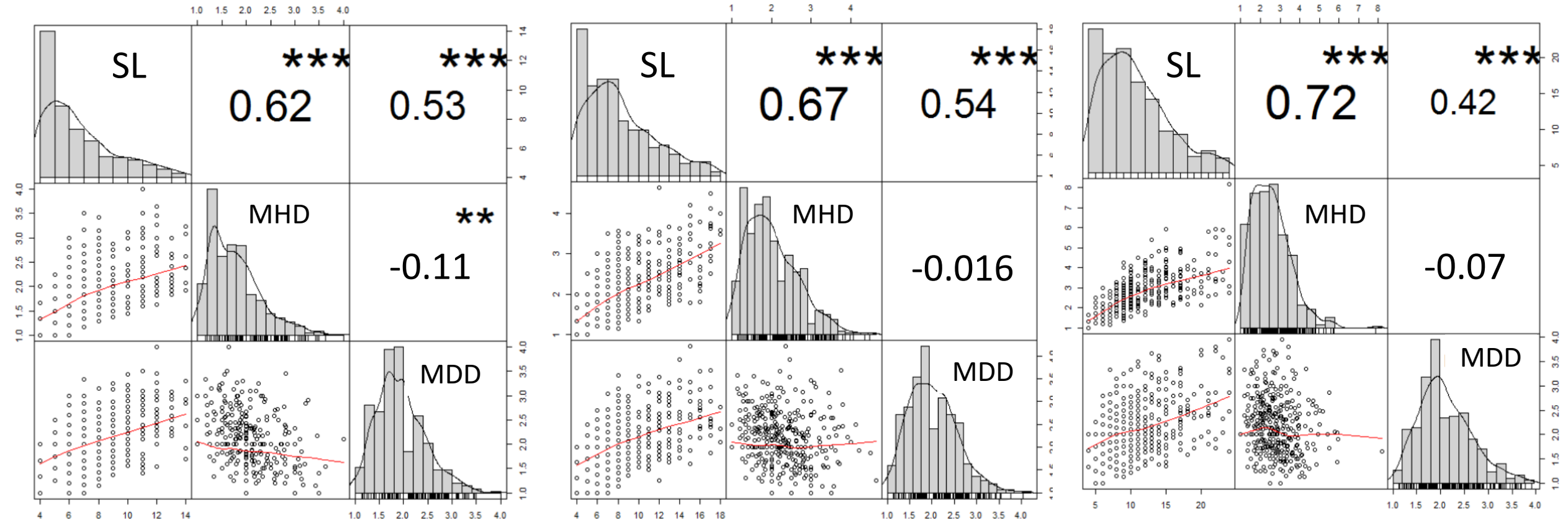
Example: MHD 5.29 > MDD 1.29 (diff=4.00)

「病院がなくなることで困難な状況に置かれる人のセーフティネットを明確にしない
まま、いきなり閉鎖をするのはいかがなものかと思う。」 *from YNU corpus, written by NS (Task 6, J017)*
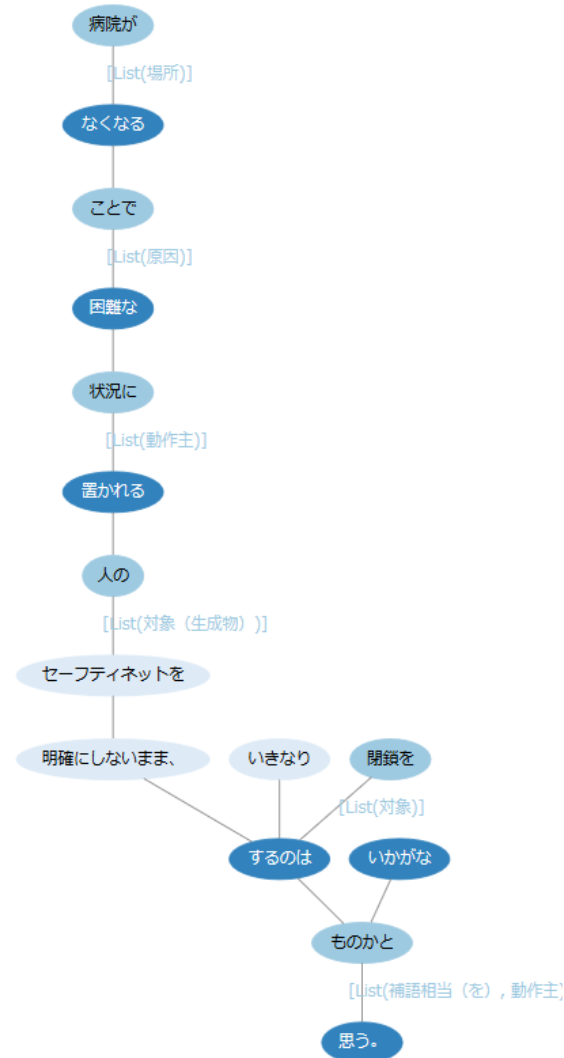
Predicate-argument structure analysis using Okayama
University ASA page
http://asap.cl.cs.okayama-u.ac.jp/asap

Example: MHD 1.18 < MDD 5.00 (diff=-3.82)

「しかし、ひこぼしは泣いてばかりいて、畑は前よりも草がたくさんはえ、牛の病気もどんどんひどくなります。」

ひこぼしは

[List(経験者)]

泣いてばかりいて、　畑は　前よりも　草が　たくさん　牛の

[List(場所)][List(場所)]

しかし、　生え、　病気も　どんどん

[List(経験者)]

ひどくなります。

Many dependent words

Summary:
comparison with previous studies

SLA studies	L1	L2	MHD	MDD
Jiang and Ouyang (2017)	Chinese	English L2	**?**	gradual increase
Komori et al. (2018 and 2019)	Chinese	Advanced Japanese L2	gradual increase (2019)	no increase (2018)
Current study	Chinese	Intermediate Japanese L2	Significant increase	no significant increase

What do MDD and MHD measure?

- Measuring different aspects of syntactic complexity

- The difference between <u>Jiang and Ouyang</u> and <u>our study</u> may be due to target language differences.   -> English   vs.   Japanese

- Jing and Liu (2015) reported Czech tends to enhance MHD whereas English prefers to increase MDD with NS data.

-Japanese is also the language with relatively free word order just like Czech. -> which may imply Japanese also enhance MHD

Different aspects of syntactic complexity?

The concept of "syntactic difficulty" consist of two factors:

1) Syntactic structure

2) Processing load

- syntactic difficulty and syntactic complexity

Language structure is not linear, however language should be produced linearly.

Therefore, language processing is affected not only structural complexity but also processing load.

7 Conclusion

This study examined and compared two syntactic analysis methods

  MDD and MHD using our original corpus data

As a result:

- Japanese learners' syntactic complexity can be measured with the MHD, but it is not as clear with the MDD

- the MHD might be a better measure to show Japanese learners' syntactic development for both intermediate and advanced learners.

- There may be a linguistic preference of MHD in Japanese.

Further studies

1) MHD of Chinese L1 English L2

2) Other combinations of L1 and L2 (**Japanese L1 English L2**)

L1 \ L2	C	J	E
C		**MDD** **MHD**	MDD ?
J	? ?		? ?
E	? ?	? ?	

References

- Jingyang Jiang and Jinghui Ouyang. 2017. Dependency distance: A new   perspective on the syntactic development in second language acquisition Comment on "Dependency distance: A new perspective on syntactic patterns in natural languages" by Haitao Liu et al. *Physics of Life Reviews* 21, 209-210.

- Yingqi Jing and Haitao Liu. 2015. Mean Hierarchical Distance Augmenting Mean Dependency Distance. *Proceedings of the Third International Conference on Dependency Linguistics,* 161-170.

- Yingqi Jing and Haitao Liu. 2016. A quantitative analysis of English hierarchical structure. *Journal of Foreign Languages,* 39, 2-11.

- Hiroyuki Kanazawa ed. 2014 *Nihongo kyoiku no tame no tasuku betsu kakikotoba kopasu* (Corpus of task-based writing for Japanese language education), Hitsuji, Tolyo.

- Saeko Komori, Masatoshi Sugiura and Wenping Li. 2018. Examining the applicability of the mean dependency distance (MDD) for SLA:A case study of Chinese learners of Japanese as a second language. *Proceedings of the 4th Asia Pacific Corpus Linguistic Conference (APCLC 2018),* 237-239.

- Saeko Komori, Masatoshi Sugiura and Wenping Li. 2019. Evaluating mean dependency distance (MDD) and mean Hierarchical distance (MHD) to measure development of Japanese syntactic complexity. *The 2019 conference of the American Association for Applied Linguistics (AAAL).*

- Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking*, CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002*, 63-69.

- Haitao Liu, Chunshan Xu and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages, *Physics of Life Reviews*, 21, 171-193.

- Lourdes Ortega. 2015. Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing,* 29, 82–94.

- Jinghui Ouyang and Jingyang Jiang. 2017. Can the probability distribution of dependency distance measure language proficiency of second language learners? *Journal of Quantitative Linguisitics,* October 2017, 1-20.

- Jinghui Ouyang and Jingyang Jian. 2018. Minimization and probability disctribution of dependency distance in the process of second language acquisition, *QUALICO,* 2018.

Acknowledgements:

Merci beaucoup!