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
2 Previous Studies on MDD and MHD
3 Research Question
4 Analysis
   4.1 Procedure
   4.2 Data
5 Results
6 Discussion
7 Conclusion
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.
## Previous studies on MDD and MHD

<table>
<thead>
<tr>
<th>Study</th>
<th>MDD/MHD</th>
<th>Language</th>
<th>NS/NNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jing and Liu (2015)</td>
<td>MDD and MHD</td>
<td>English and Czech</td>
<td>NS</td>
</tr>
<tr>
<td>Liu et al. (2017)</td>
<td>MDD</td>
<td>20 natural languages</td>
<td>NS</td>
</tr>
<tr>
<td>Ouyang and Jiang (2017)</td>
<td>MDD</td>
<td>English as a second language</td>
<td>NNS</td>
</tr>
<tr>
<td>Komori et al. (2018, 2019)</td>
<td>MDD and MHD</td>
<td>Japanese</td>
<td>NS/NNS</td>
</tr>
</tbody>
</table>
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.
Table 3: Correlations between SL, MDD, and MHD

<table>
<thead>
<tr>
<th>Lang</th>
<th>X-Y</th>
<th>Cor</th>
<th>p</th>
<th>k</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>en</td>
<td>SL-MDD</td>
<td>0.54</td>
<td>&lt;0.01</td>
<td>0.03</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>SL-MHD</td>
<td>0.74</td>
<td>&lt;0.01</td>
<td>0.09</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>MDD-MHD</td>
<td>0.19</td>
<td>&lt;0.01</td>
<td>0.41</td>
<td>0.04</td>
</tr>
<tr>
<td>cs</td>
<td>SL-MDD</td>
<td>0.42</td>
<td>&lt;0.01</td>
<td>0.02</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>SL-MHD</td>
<td>0.74</td>
<td>&lt;0.01</td>
<td>0.12</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>MDD-MHD</td>
<td>0.11</td>
<td>&lt;0.01</td>
<td>0.36</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 3 and Figure 7 of Jing and Liu (2015)
2.2 Liu et al. (2017)
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’s 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 ($p=0.782>0.005$).
2.4 Advanced Japanese learners’ study (Komori et al., 2018 and 2019)

<table>
<thead>
<tr>
<th>YNU</th>
<th>CL</th>
<th>CM</th>
<th>CH</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD</td>
<td>2.16</td>
<td>2.08</td>
<td>2.16</td>
<td>2.07</td>
</tr>
<tr>
<td>MHD</td>
<td>1.75</td>
<td>1.84</td>
<td>1.98</td>
<td>1.97</td>
</tr>
</tbody>
</table>

**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
Example sentence:
“Kono tabiwa oukagaisitai kotoga ari, meeruwo okuraseteitadakimasita.”
(I sent an email because I have something that I would like to ask you.)

from YNU corpus, written by NS (Task 1, J001)

MDD = (1+5+1+1+2+1) ÷ (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

<table>
<thead>
<tr>
<th>Group</th>
<th>Participants</th>
<th>Sentences</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2 (second year university learners)</td>
<td>38</td>
<td>721</td>
<td>10,296</td>
</tr>
<tr>
<td>C3 (third year university learners)</td>
<td>33</td>
<td>605</td>
<td>11,786</td>
</tr>
<tr>
<td>JP (Japanese university students)</td>
<td>35</td>
<td>463</td>
<td>12,495</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Number of Sentences</th>
<th>Sentence Length (SL) (Min, Max)</th>
<th>Median MDD (Min, Max)</th>
<th>Median MHD (Min, Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>592</td>
<td>6 (4, 4)</td>
<td>1.91 (1.00, 4.00)</td>
<td>1.67 (1.00, 4.00)</td>
</tr>
<tr>
<td>C3</td>
<td>547</td>
<td>8 (4, 18)</td>
<td>2.00 (1.00, 4.21)</td>
<td>2.00 (1.00, 4.64)</td>
</tr>
<tr>
<td>JP</td>
<td>429</td>
<td>10 (4, 24)</td>
<td>2.00 (1.00, 3.96)</td>
<td>2.50 (1.00, 8.17)</td>
</tr>
</tbody>
</table>
MDD and MHD
MDD

C2

C3

JP
<table>
<thead>
<tr>
<th></th>
<th>MDD</th>
<th></th>
<th>MHD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BM</td>
<td>p</td>
<td>Cliff’s delta</td>
</tr>
<tr>
<td>C2 v. C3</td>
<td>3.88</td>
<td>.0001</td>
<td>.13 (negligible)</td>
</tr>
<tr>
<td>C3 v. JP</td>
<td>1.04</td>
<td>.2988</td>
<td>.04 (negligible)</td>
</tr>
<tr>
<td>C2 v. JP</td>
<td>4.86</td>
<td>&lt;.0001</td>
<td>.17 (small)</td>
</tr>
</tbody>
</table>
Correlations between SL, MDD and MHD

C2

SL 0.62 0.53
MHD -0.11

C3

SL 0.67 0.54
MHD -0.016

JP

SL 0.72 0.42
MHD -0.07

MDD
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。」

from YNU corpus, written by NS (Task 12, J029)
Summary: comparison with previous studies

<table>
<thead>
<tr>
<th>SLA studies</th>
<th>L1</th>
<th>L2</th>
<th>MHD</th>
<th>MDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jiang and Ouyang (2017)</td>
<td>Chinese</td>
<td>English L2</td>
<td>?</td>
<td>gradual increase</td>
</tr>
<tr>
<td>Current study</td>
<td>Chinese</td>
<td>Intermediate Japanese L2</td>
<td>Significant increase</td>
<td>no significant increase</td>
</tr>
</tbody>
</table>
What do MDD and MHD measure?

• Measuring different aspects of syntactic complexity
• The difference between Jiang and Ouyang and our study 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)

<table>
<thead>
<tr>
<th>L1</th>
<th>C</th>
<th>J</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td></td>
<td>MDD MHD</td>
<td>MDD ?</td>
</tr>
<tr>
<td>J</td>
<td>?</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>E</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
References


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

• 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).*


Acknowledgements:

• This work was supported by
  JSPS KAKENHI Grant Number JP19K00749 and
  JSPS KAKENHI Grant Number 16H03444
Merci beaucoup!