Reading demands & language proficiency in foreign language learning
An exploration of linguistic complexity modeling

Zarah Weiss   Detmar Meurers
University of Tübingen
Germany

Text Comprehension: What We Know About the Dance Between Reader, Text, and Task in Reading Comprehension
AERA Symposium, April 12., 2021
Language input fosters foreign language (FL) learning
  ▶ Challenging input at i+1 (Krashen 1981)
  ▶ Zone of Proximal Development (ZPD; Lantolf et al. 2015)

Identification of FL reading materials requires
  1. analysis of FL proficiency of language learner
  2. analysis of FL reading materials
  3. match of proficiency and reading level

We use linguistic complexity analysis to link FL writing complexity and FL reading materials:
  ▶ Is the input appropriately challenging (“+1”)?
    Empirical focus here: German as a FL
Automatic analysis of linguistic complexity

- Complexity, Accuracy, and Fluency used to describe language performance in SLA (Housen et al. 2019)
  - complex language is elaborate and varied (Ellis 2003)

- Characterization of proficiency and development (Housen et al. 2019; Ortega 2012; Crossley & McNamara 2014)

- For German, we analyze 489 complexity feature:
  - covering syntax, lexicon, morphology, cohesion and human language processing and use measures
  - successfully used for
    - proficiency assessment: Weiss & Meurers (2019a,b, 2021)
First analysis: Readability

▶ We collected 1,446 graded GFL reading texts
  ▶ by a leading European publisher of FL reading materials
▶ Contains easy, medium, and advanced texts
  ▶ linked to CEFR levels A2, B1/B2, and C1 by publisher
▶ Automatic prediction of reading levels A2, B1/B2, and C1
  ▶ Trained ordinal random forest with 10-fold cross-validation
  ▶ Accuracy = 89.0% in 3-way classification
Second analysis: Proficiency

- **MERLIN**: cross-sectional German as a FL data (N=1,033; Wisniewski et al. 2013)
  - elicited in official standardized language certification tests
- Learner writing rated by two experts on Common European Framework of Reference (CEFR) scale
- **⇒** Automatic CEFR classification (A1, A2, B1, B2, C):
  - trained ordinal random forest with 10-fold cross-validation
  - accuracy for 5-way classification = 70.0%
Zooming in on relevant features across domains

- 373 features are informative (info gain) for both data sets
- Complexity feature values for reading texts are higher than those for learner productions for

<table>
<thead>
<tr>
<th></th>
<th>A2</th>
<th>B1/B2</th>
<th>C1</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>38%</td>
<td>25%</td>
<td>7%</td>
<td>60</td>
</tr>
<tr>
<td>Lexicon</td>
<td>82%</td>
<td>67%</td>
<td>56%</td>
<td>95</td>
</tr>
<tr>
<td>Morphology</td>
<td>88%</td>
<td>73%</td>
<td>35%</td>
<td>48</td>
</tr>
<tr>
<td>Cohesion</td>
<td>97%</td>
<td>86%</td>
<td>72%</td>
<td>36</td>
</tr>
<tr>
<td>Language use</td>
<td>91%</td>
<td>71%</td>
<td>37%</td>
<td>109</td>
</tr>
</tbody>
</table>

- We zoom in on features to explore where an adaptive “+1” relation can be observed in the different domains
  - syntax, lexicon, morphology, cohesion
Systematic increase syntactic elaborateness in both
but values for reading texts are systematically lower
⇒ no syntactic challenge in reading texts
Lexical complexity

- Systematic increase in lexical variation for both
- MTLD and corrected verb variation higher for reading texts
  ⇒ texts offer challenge for learners in lexical domain
Morphological complexity

- Systematic increase in morphological variation for both
  - Morphological complexity index measures POS-specific TTR for morphological units (Brezina & Pallotti 2019)
  - Nominal and verbal MCI higher for A and B1 levels
    ⇒ Texts offer challenge for beginning/intermediate
Cohesion

- Reading texts systematically use more connectives and higher implicit cohesion
  ⇒ Texts offer challenge with respect to text cohesion
Summary and Outlook

- Leveled reading material generally more complex than learner production in most domains but syntax.
  - most prominent for beginning and intermediate levels
  - to be clarified: How much of this is due to gap between active and passive language competence?

⇒ Which level of challenge in which domain best fosters foreign language learning for which kind of learners?
  - More research needed on the nature of the adaptivity best fostering foreign language learning.
  - Linguistic complexity analysis provides fine-grained instrument to carry out such research.
References


