Automated assessment of lexical diversity and n-grams in the Pearson Test of English Academic Language Testing Forum, Bristol University, 16-18 November 2012

Jeanine Treffers-Daller | Patrick Parslow | Shirley Williams

Research questions:

1. Which range of lexical diversity scores are typically found in essays written by learners of English at B1, B2, C1 and C2 of the CEFR?
2. Which n-grams can discriminate between different levels of the CEFR?
3. To what extent do lexical diversity scores and n-gram usage correlate with scores obtained on the PTE Academic?

Data:

<table>
<thead>
<tr>
<th>CEFR level</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>29</td>
</tr>
</tbody>
</table>

- PTE Academic Scores: overall score PTE Academic, Writing Score, Vocabulary Score (range: 10-90 points for each score).
- Lexical diversity scores: the range of vocabulary displayed in a text (Malvern et al 2004)
  - Traditional measure: Type/token ratio (Johnson 1944; Templin 1957)
  - Index of Guiraud: types/√tokens (Guiraud 1954)
  - D: model of the falling TTR curve (Malvern et al 2004; Treffers-Daller & Parslow in press)

Results 1: LD and the CEFR

\[ F(3,175)=6.9, p <.001 \]
\[ \text{Eta squared: .11} \]

Group B1 significantly different from groups B2, C1 and C2. Other intergroup differences n.s.

Results 2: N-grams

Results 3: correlations between PTE scores and LD scores

Conclusions:

- There are significant differences in lexical diversity shown between CEFR grades. The Index of Guiraud appears to correlate slightly more strongly with CEFR grades than D. N-gram use exhibits differing correlations by grade and value of n. The overlap between CEFR grades for each of these metrics is insufficient to provide discriminatory power, but it remains feasible that combinations of metrics can provide adequate classification.

Further research:

We will compare the usage of n-grams in academic texts from the PICAE corpus (Ackermann et al. 2010) with those found in the students’ test data and investigate how the analysis of lexical diversity and n-grams can contribute to the development of criterion features that distinguish between levels of the CEFR.

Bibliography


Acknowledgements

We are very grateful to Pearson for their sponsorship of the research and to Kirsten Ackermann and Veronica Benigno from Pearson for their advice in the handling of the data.

Contact information

- Department of Education and School of Systems Engineering, University of Reading
- Email j.c.treffers-daller@reading.ac.uk | p.parslow@reading.ac.uk | shirley.williams@reading.ac.uk