New Approaches to ICE texts: a critical, empirical comparison of Academic Writing in ICE-EA and beyond

Josef Schmied
English Language & Linguistics
Chemnitz University of Technology
www.tu-chemnitz.de/phil/english/schmied
josef.schmied@tu-chemnitz.de

Aims

Databases/Corpora
Concepts & Methods
Comparison: ICE
Comparison: Nordic
Outlook

1. Databases/Corpora

- ICE - East Africa (1990-96):
  a stratified corpus of English as a Second Language: 500 text-types of 2000 words each
  but ICE-K/ICE-T adapted → 1.5 Million words
  informational learned + popular 0.16 Million words

- Nordic Journal of African Studies:
  English as an Academic lingua franca, mainly for African scholars
  37 articles with 0.25 Million words

Appendix 6: List of written texts from Tanzania (word count)

<table>
<thead>
<tr>
<th>Informational: Learned</th>
<th>Humanities</th>
<th>Natural Science</th>
<th>Technology/Agriculture/Environmental dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W2A001T - W2A010T</td>
<td>20.172</td>
<td>20.114</td>
</tr>
<tr>
<td></td>
<td>W2B010T - W2B030T</td>
<td>20.151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W2B031T - W2B032T</td>
<td>6.542</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W2B033T - W2B034T</td>
<td>20.061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>W2B035T - W2B036T</td>
<td>13.788</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80.752</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Informational: Popular

<table>
<thead>
<tr>
<th>Humanities</th>
<th>Natural Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>W2B011T - W2B020T</td>
<td>20.133</td>
</tr>
<tr>
<td>W2B021T - W2B022T</td>
<td>20.223</td>
</tr>
<tr>
<td>W2B023T - W2B024T</td>
<td>6.542</td>
</tr>
<tr>
<td>W2B025T - W2B026T</td>
<td>20.061</td>
</tr>
<tr>
<td>80.752</td>
<td></td>
</tr>
</tbody>
</table>

Informational: Reportage

<table>
<thead>
<tr>
<th>Splash</th>
<th>Reportage/Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>W2A010T - W2A020T</td>
<td>20.018</td>
</tr>
<tr>
<td>W2A011T - W2A022T</td>
<td>20.133</td>
</tr>
<tr>
<td>41.157</td>
<td></td>
</tr>
</tbody>
</table>

Institutional/administrative/regulatory

<table>
<thead>
<tr>
<th>Persuasive</th>
<th>Institutional</th>
</tr>
</thead>
<tbody>
<tr>
<td>W2B011T - W2B020T</td>
<td>20.078</td>
</tr>
<tr>
<td>W2B011T - W2B022T</td>
<td>20.125</td>
</tr>
<tr>
<td>41.201</td>
<td></td>
</tr>
</tbody>
</table>
2. Concepts & Methods

**Automatic corpus processing:**
POS tagging using Penn Treebank + treetagger

**ComplexAna:** Complexity Analyser
morphosyntactic (type/token) and
semantic (unknown words in WordNet)
with flexible parameter weight

**Human corpus analysis:**
adjuncts: linking, modal, evaluative, domain-specific

**Searches** with AntConc for 181 adverbs

### Table: Modal Adjuncts in Huddleston/Pullum 2006:768

<table>
<thead>
<tr>
<th>Strong</th>
<th>Assuredly</th>
<th>Certainly</th>
<th>Clearly</th>
<th>Definitely</th>
<th>Incontestably</th>
<th>Ineluctably</th>
<th>Indubitably</th>
<th>Unarguably</th>
<th>Unavoidably</th>
<th>Unassumingly</th>
<th>Unquestionably</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indubitably</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ineluctably</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inassumingly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unquestionably</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unarguably</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unavoidably</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Evaluative Adjuncts in Huddleston/Pullum 2006:771

- Accordingly
- Actually
- Additionally
- Admittedly
- Alternately
- Analytically
- Apparently
- Astonishingly
- Asymmetrically
- Asymptotically
- Automatically
- Autonomously
- Basically
- Briefly
- Carefully
- Clearly
- Certainly
- Clearly
- Definitely
- Incontestably
- Ineluctably
- Indubitably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Unquestionably
- Unarguably
- Unavoidably
- Unassumingly
- Una
3. Comparison: Ke and Tz in ICE-EA

Hypotheses:
- Complexity ICE-KE > ICE-TZ
- Clause linking by adjuncts ICE-KE > ICE-TZ

including a few surprises

3.2. Clause linkers ICE-EA spoken

Hypo supported: more clause linkers in KE

4. Comparison: Ke and Tz in Nordic

Hypotheses:
- Complexity ICE-KE > ICE-TZ
- Clause linking by adjuncts ICE-KE > ICE-TZ

including a few surprises
### 4.1. Complexity of KE/TZ articles in Nordic Journal Corpus

<table>
<thead>
<tr>
<th>Complex/Corpora</th>
<th>KE01h</th>
<th>KE02h</th>
<th>T201h</th>
<th>T202h</th>
<th>CM all</th>
<th>UK01h</th>
<th>mean ± 2SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tokens</td>
<td>3,530</td>
<td>3,000</td>
<td>2,666</td>
<td>3,044</td>
<td>4,869</td>
<td>2,500</td>
<td>3,570 ± 219</td>
</tr>
<tr>
<td>Max. words per sentence</td>
<td>92</td>
<td>66</td>
<td>28</td>
<td>8</td>
<td>92</td>
<td>22</td>
<td>69 ± 10</td>
</tr>
<tr>
<td>Mean words per sentence</td>
<td>13.99</td>
<td>21.87</td>
<td>18.35</td>
<td>14.95</td>
<td>23.15</td>
<td>26.85</td>
<td>20.89 ± 28.19</td>
</tr>
<tr>
<td>Nouns</td>
<td>3,099</td>
<td>1,702</td>
<td>1,202</td>
<td>1,502</td>
<td>2,262.43</td>
<td>1,898</td>
<td>2,526.33 ± 219.56</td>
</tr>
<tr>
<td>Mean length noun</td>
<td>402</td>
<td>616</td>
<td>794</td>
<td>978</td>
<td>908.95</td>
<td>1,808</td>
<td>907.76 ± 82.60</td>
</tr>
<tr>
<td>Max. length noun</td>
<td>68.41</td>
<td>89.59</td>
<td>89.02</td>
<td>72.22</td>
<td>76.76</td>
<td>79.86</td>
<td>78.96 ± 73.02</td>
</tr>
<tr>
<td>Nouns not in frequency list (%)</td>
<td>36.0</td>
<td>28.4</td>
<td>44.7</td>
<td>38.4</td>
<td>41.6</td>
<td>45.1</td>
<td>41.5 ± 5.80</td>
</tr>
<tr>
<td>Max. length noun</td>
<td>42</td>
<td>42</td>
<td>23</td>
<td>21</td>
<td>42</td>
<td>42</td>
<td>23.0 ± 4.20</td>
</tr>
<tr>
<td>Mean length noun</td>
<td>8.71</td>
<td>7.34</td>
<td>7.07</td>
<td>7.11</td>
<td>7.18</td>
<td>6.30</td>
<td>6.95 ± 0.05</td>
</tr>
<tr>
<td>Max. content in a sentence</td>
<td>175</td>
<td>239</td>
<td>187</td>
<td>170</td>
<td>197.06</td>
<td>206</td>
<td>198.28 ± 22.22</td>
</tr>
<tr>
<td>Max. content in a sentence</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>10</td>
<td>10.00</td>
<td>9</td>
<td>9.99 ± 0.99</td>
</tr>
<tr>
<td>Nouns not in frequency list (%)</td>
<td>26</td>
<td>28</td>
<td>44</td>
<td>38</td>
<td>41</td>
<td>45</td>
<td>41.5 ± 5.80</td>
</tr>
<tr>
<td>Max. length noun</td>
<td>8.34</td>
<td>8.61</td>
<td>8.39</td>
<td>8.55</td>
<td>8.51</td>
<td>8.51</td>
<td>8.51 ± 0.51</td>
</tr>
</tbody>
</table>

### 4.2. Clause Linkers

#### Conjunctions

<table>
<thead>
<tr>
<th>Clause Linker</th>
<th>KE01h</th>
<th>KE02h</th>
<th>T201h</th>
<th>T202h</th>
<th>CM all</th>
<th>UK01h</th>
<th>mean ± 2SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>But</td>
<td>24</td>
<td>8</td>
<td>6</td>
<td>30</td>
<td>15</td>
<td>6</td>
<td>19.1 ± 3.5</td>
</tr>
<tr>
<td>Although</td>
<td>12</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3.5 ± 1.4</td>
</tr>
<tr>
<td>While</td>
<td>48</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>13.8 ± 7.6</td>
</tr>
<tr>
<td>If</td>
<td>12</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>14.2 ± 4.2</td>
</tr>
<tr>
<td>Whether</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>4.2 ± 4.2</td>
</tr>
<tr>
<td>Because</td>
<td>4</td>
<td>4</td>
<td>50</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>11.8 ± 4.8</td>
</tr>
<tr>
<td>In order to</td>
<td>4</td>
<td>4</td>
<td>40</td>
<td>50</td>
<td>2</td>
<td>4</td>
<td>4.0 ± 4.0</td>
</tr>
<tr>
<td>Since</td>
<td>18</td>
<td>30</td>
<td>10</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>6.4 ± 7.6</td>
</tr>
<tr>
<td>Sum conjuncts</td>
<td>126</td>
<td>58</td>
<td>158</td>
<td>58</td>
<td>43</td>
<td>38</td>
<td>67.8 ± 38.7</td>
</tr>
</tbody>
</table>

### 5. Outlook

- **expand the linguistic perspective to comparative discourse and text features** (e.g. complexity, cohesion, coherence)
- **expand the data base to achieve significance**: NORDIC Journal and ICEweb-Ka-Tz-Ug (>2 Mill words in 4 domains)
- **compare varieties intra- and internationally**, incl. diachronic and text-type variation
- **writer perspective**: construct sequences of text features in (Non-Native?) Englishes, e.g. from formal explicit cohesion to semantic implicit cohesion
- **reader perspective**: re-consider contextualisation and adaptation