GRAMMATICAL AND LEXICAL COMPLEXITY

In bilingual speakers (L2 speakers & L1 attriters) beyond the advanced stage

Cornelia Lahmann & Rasmus Steinkrauss
VUB, Brussels
19-12-2014
1. Theoretical background
   • Multi-dimensional complexity in the CAF setting
     • Grammatical
     • Lexical

2. Research Question: Complexity beyond the advanced stage?

3. Study
   • Method
     • Speakers: L2 users, L1 attriters
     • Data: syntactic, morphological & lexical measures
     • Hierarchical cluster & principal component analysis

   • Results
   • Discussion

4. Conclusion
Multi-dimensional complexity in the CAF setting

- Analytic view (Bulté & Housen)
  - Complexity in absolute terms: the number of elements and their relationship (structure and system complexity)
    ➔ To measure the (L2) learner’s linguistic system (Bulté & Housen)

- Holistic view (Norris & Ortega; Byrnes)
  - Interrelatedness of complexity, accuracy, and fluency (CAF)
  - Interrelatedness of linguistic domains

- Simple view (Pallotti)
  - Complexity in structural terms
    ➔ To assess text complexity

- The CAF setting:
  - *Mostly* instructed, school/university settings
  - *Rarely* naturalistic, immersion settings
Grammatical complexity (analytic view)

• Syntactic complexity (functional)
  • Sentence, clausal, phrasal complexity
    → Measures of length and subordination

• Morphological complexity (formal)
  • Inflectional and derivational complexity
    → Number of exponents (the forms taken by lexemes to express grammatical categories and functions)

[Bulté & Housen 2012, 2014; Pallotti 2014]
Lexical complexity (analytic view)

- Diversity
  - TTRs

- Density
  - Content words in relation to function words

- Sophistication
  - Frequency-based lexical measures
  - Semantic measures

- Compositionality
  - Length of words in terms of morphemes and syllables

Summary

Complexity usually assessed…
- for beginning to advanced learners
- in instructed settings
- by means of writing assessments
  …with measures of…
- length and subordination
- diversity
Complexity beyond the advanced stage

Complexity usually assessed…
• for beginning to advanced learners
• in instructed settings
• by means of writing assessments
…with measures of…
• length and subordination
• diversity

Needed:
⇒ very advanced
⇒ naturalistic setting
⇒ spontaneous speech
⇒ more / different measures
Complexity beyond the advanced stage

Complexity usually assessed...
- for beginning to advanced learners
- in instructed settings
- by means of writing assessments
...with measures of...
- length and subordination
- diversity

Needed:
- very advanced
- naturalistic setting
- spontaneous speech
- more / different measures

To what extent does the analytic, multi-dimensional view of complexity hold in a corpus of spontaneous speech produced by two groups of “bilinguals” beyond the advanced stage?
Data

• Free speech: oral history interviews (~30 mins)
  • Including anecdotes, commentaries, descriptions
• German Jewish immigrants
• Escaping from Germany in 1938/39 before WW II
  • on their own: Kindertransport (children's transport)
  • or with their family
• To UK, USA, Australia
# Speakers

<table>
<thead>
<tr>
<th></th>
<th>M (SD)</th>
<th>Range/Categories</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2S (n = 102)</td>
<td>L1A (n = 73)</td>
<td>L2S (n = 102)</td>
<td>L1A (n = 73)</td>
</tr>
<tr>
<td>AE</td>
<td>12.15 (2.67)</td>
<td>13.8 (2.08)</td>
<td>7-17</td>
<td>7-17</td>
</tr>
<tr>
<td>LoR</td>
<td>61.33 (6.12)</td>
<td>60.3 (5.5)</td>
<td>41-73</td>
<td>46-70</td>
</tr>
<tr>
<td>Aal</td>
<td>73.59 (6.97)</td>
<td>74.08 (6.14)</td>
<td>57-87</td>
<td>58-85</td>
</tr>
<tr>
<td>L1 Exp</td>
<td>4.34 (1.44)</td>
<td>4.62 (1.49)</td>
<td>1-7</td>
<td>1-7</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td>M: 42 F: 60</td>
<td>M: 38 F: 35</td>
<td></td>
</tr>
<tr>
<td>Edu</td>
<td></td>
<td>Low: 11 Mid: 34 High: 46</td>
<td>Low: 9 Mid: 22 High: 31</td>
<td></td>
</tr>
</tbody>
</table>

*Note: L2S = L2 speakers, L1A = L1 attriters, AE = Age at emigration, LoR = length of residence, Aal = Age at interview, L1 Exp = Continued L1 exposure after emigration, Edu = Level of education*
Data

- **Syntactic (complexity) tagging**

*XYZ: so she was away (..) for quite some time.*

%xcsy: UTT|AS|MC:8|SUBJ:1|V:1:0|NP:2:4|

*XYZ: vielleicht werden sie dich gehen lassen.*

%xcsy: UTT|AS|MC:6|SUBJ:1|V:1:2|NP:2:2|KV|

- **Morphsyntactic tagging**
  - MOR (CLAN, CHILDES, MacWhinney, 2000)
  - Tree Tagger software (University of Tübingen, H. Schmid, 1995)
## Measures

<table>
<thead>
<tr>
<th>Linguistic aspect</th>
<th>Level</th>
<th>Statistical construct</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Syntactic</strong></td>
<td>Sentence &amp; clause</td>
<td>Words per utterance, words per AS-unit, words per clause</td>
</tr>
<tr>
<td></td>
<td>Sub-clause</td>
<td>Dependent clauses per AS-unit, <strong>non-finite adverbial DCs per AS-unit</strong>, object-subject relative clause ratio</td>
</tr>
<tr>
<td></td>
<td>Phrase</td>
<td>Words per NP, words per SUBJ</td>
</tr>
<tr>
<td><strong>Morphological</strong></td>
<td></td>
<td>Passives, <strong>subjunctives, complex VPs</strong> per clause</td>
</tr>
<tr>
<td><strong>Lexical</strong></td>
<td>Diversity</td>
<td>TTR, Guiraud</td>
</tr>
<tr>
<td></td>
<td>Sophistication</td>
<td>Frequency Bands for content words, <strong>BNC spoken frequency</strong>, BNC spoken trigram frequency, Hypernymy, Polysemy, concreteness, imageability</td>
</tr>
</tbody>
</table>
Hierarchical cluster analysis (HCA)

• Technical details
  • *R* software, *hclust* package
  • Iterative, agglomerative algorithm joins least dissimilar clusters at each stage
  • Ward’s minimum variance method
  • Outputs dendrogram illustrating the hierarchical cluster structure

• Analysis:
  • Initially each measure is assigned to its own cluster, after which the next higher cluster is built
  • Finally there is a single cluster with several subclusters
Principal component analysis (PCA)

• Technical details
  • R software, GPArotation & psych packages
  • Multivariate technique to represent original variables as a set of principal components
  • Bartlett’s test and Kaiser-Meyer-Olkin (KMO) for sampling adequacy
  • Outputs a table of factor loadings

• Analysis:
  • Initially there is a correlation matrix with all measures representing one principal component (cluster)
  • Finally there is a matrix with several principal components (clusters), onto which individual measures load
Results: L2 speakers (Grammatical complexity - HCA)

1: Theoretical background
2: Research Question
3: Study
4: Conclusion
## Results: L2 speakers (Grammatical complexity - PCA)

<table>
<thead>
<tr>
<th>Item/measure</th>
<th>Oblique rotated factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>WD_SUB</td>
<td>0.83</td>
</tr>
<tr>
<td>WD_CLA</td>
<td>0.81</td>
</tr>
<tr>
<td>WD_NP</td>
<td>0.77</td>
</tr>
<tr>
<td>PASS_CLA</td>
<td>0.66</td>
</tr>
<tr>
<td>WD_AS</td>
<td>0.59</td>
</tr>
<tr>
<td>WD_U TT</td>
<td>0.56</td>
</tr>
<tr>
<td>NONFDC_AS</td>
<td>-0.13</td>
</tr>
<tr>
<td>DC_AS</td>
<td>0.22</td>
</tr>
<tr>
<td>OBJREL_SUBREL</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

| Eigenvalues | 3.32 | 2.24 | 1.04 |
| % of variance | 36.89 | 24.89 | 11.56 |

1: Theoretical background 2: Research Question 3: Study 4: Conclusion
Results: L1 attriters (Grammatical complexity - HCA)

1: Theoretical background 2: Research Question 3: Study 4: Conclusion
### Results: L1 attriters (Grammatical complexity - PCA)

<table>
<thead>
<tr>
<th>Item/measure</th>
<th>Oblique rotation factor loadings</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Item/measure</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>WD_NP</td>
<td>.91</td>
<td>-.09</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>WD_CLA</td>
<td>.90</td>
<td>.10</td>
<td>.08</td>
<td></td>
</tr>
<tr>
<td>WD_AS</td>
<td>.86</td>
<td>.25</td>
<td>-.07</td>
<td></td>
</tr>
<tr>
<td>WD_SUBJ</td>
<td>.76</td>
<td>-.37</td>
<td>-.23</td>
<td></td>
</tr>
<tr>
<td>DC_AS</td>
<td>.58</td>
<td>.35</td>
<td>-.24</td>
<td></td>
</tr>
<tr>
<td>WD_U TT</td>
<td>.43</td>
<td>-.01</td>
<td>-.32</td>
<td></td>
</tr>
<tr>
<td>KV_CLA</td>
<td>-.07</td>
<td>.79</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>KONJ_CLA</td>
<td>.21</td>
<td>.73</td>
<td>.12</td>
<td></td>
</tr>
<tr>
<td>PASS_CLA</td>
<td>.10</td>
<td>.63</td>
<td>-.35</td>
<td></td>
</tr>
<tr>
<td>OBJ_SUBJ</td>
<td>.05</td>
<td>.02</td>
<td>.91</td>
<td></td>
</tr>
<tr>
<td>Eigenvalues</td>
<td>3.67</td>
<td>1.97</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>% of variance</td>
<td>36.7</td>
<td>19.7</td>
<td>12.8</td>
<td></td>
</tr>
</tbody>
</table>
Discussion Grammatical Complexity

- Dimensions
  - Length (at all levels)
  - Subordination
  - Morphology
- Statistical artefact
- Cross-linguistic differences
- OBJ_SUB clusters separately, indicating that it measures something else
- WD_UTT poor measure
Results: L2 speakers (Lexical complexity - HCA)

1: Theoretical background 2: Research Question 3: Study 4: Conclusion
## Results: L2 speakers (Lexical complexity - PCA)

<table>
<thead>
<tr>
<th>Item/measure</th>
<th>Oblique rotated factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TTR</td>
<td>0.92</td>
</tr>
<tr>
<td>F5 Ratio</td>
<td>0.91</td>
</tr>
<tr>
<td>Concreteness</td>
<td>-0.07</td>
</tr>
<tr>
<td>Imageability</td>
<td>-0.02</td>
</tr>
<tr>
<td>Hypernymy</td>
<td>0.49</td>
</tr>
<tr>
<td>Polysemy</td>
<td>0.10</td>
</tr>
<tr>
<td>F1 Ratio</td>
<td>-0.48</td>
</tr>
<tr>
<td>BNC Spoken Trigram Frequency</td>
<td>-0.44</td>
</tr>
<tr>
<td>BNC Spoken Frequency</td>
<td>-0.43</td>
</tr>
<tr>
<td>Eigenvalues</td>
<td>2.86</td>
</tr>
<tr>
<td>% of variance</td>
<td>31.78</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.91</td>
</tr>
</tbody>
</table>
Results: L1 attriters (Lexical complexity)
Discussion Lexical Complexity

• Diversity measures: TTR/Guiraud, F5 Ratio

• Sophistication measures capturing:
  • Frequency: Polysemy, F1 Ratio, BNC-based ratios
  • Abstractness: hypernymy, concreteness, imageability
Conclusion

• Multi-dimensionality of complexity construct
• Cross-linguistic differences
• Careful choice of lexical measures
• Future research
  • Further exploration of sophistication: different measures of sophistication
    • Polysemy
    • Hypernymy
    • Concreteness
    • Imageability
    • etc.
Thank you

• To Prof. A. Housen & Dr. B. Bulté

• This research would not have been possible without the support of the:
  • NWO (Grant No. 360-70-420)
  &
  • USC Shoah Foundation Institute (Los Angeles)
  • Werkstatt der Erinnerung (Hamburg)
  • Alte Synagoge (Essen)
  • Fortunoff Video Archive for Holocaust Testimonies (Yale University Library)
  • Tauber Holocaust Library and Education Program (San Francisco)
  • United States Holocaust Memorial Museum (Washington, D.C.)
  • Association of Jewish Refugees (London)
References

