Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study

Nina Vyatkina
University of Kansas

Hagen Hirschmann,
Felix Golcher
Humboldt-Universität zu Berlin

Kolloquium Korpuslinguistik
October 29, 2014
HU Berlin
Study background

• Submitted to the Special Issue of Journal of Second Language Writing “New Developments in the Study of L2 Writing Complexity”
• To be presented at the U of Groningen on Nov. 7, 2014
• Substantial revision of the AAAL 2013 presentation at the L2 Complexity colloquium
Research questions

1. (How) does the size and range of the syntactic modification system in the writing of beginning learners of German change over two years of instructed language study?

2. What are the group developmental trends in the frequency of different modifier classes and (how) are they modulated by individual differences and observation points?
L2 Complexity

• **CAF** Measures of learner progress and proficiency indicators employed in SLA since 1980s (Larsen-Freeman, 1983; Skehan, 1989)
  – **Complexity:**
    • the extent to which the language produced in performing a task is elaborate and varied (Ellis, 2003)
    • the range of forms that surface in language production and the degree of sophistication of such forms (Ortega, 2003)
  – **Accuracy:** ability to produce error-free speech or writing
  – **Fluency:** rapidity of L2 production
Problems with the notion of L2 complexity

• **Complexity** as a basic and valid
  - **descriptor** of language **performance**
  - **indicator** of language **proficiency**
  - **index** of language **progress** and **development**


=> Claim: L2 systems & L2 production become more ‘complex’ over time …
  • Increased range, breadth, depth, sophistication, compositionality, etc. of lexical and morpho-syntactic elements

⇒ BUT: is ‘more complex’ = ‘more developed’ = ‘more complex’?
  -> empirical question!

Pallotti (2009), Bulté & Housen (2012)

Adopted from Bulté & Housen (2013)
Risk of circular reasoning

Adopted from Bulté & Housen (2013)
L2 complexity: a *complex* notion

Adopted from Bulté & Housen (2012, 2013)
L2 complexity in this study
according to Bulté & Housen (2012)

• **Scope**: absolute complexity
  – number of constituent components and relationships between components

• **Theoretical level**: systemic aspects
  – elaboration, size, range, variation, ‘breadth’ of L2 grammar

• **Observational level**: grammatical diversity and sophistication
  – at the sentential, clausal, and phrasal level

• **Operational level**: 
  – syntactic modification: optional elements that attach to the heads of noun phrases, verb phrases, or sentences
This study

Adopted from Bulté & Housen (2012, 2013)
L2 complexity: research syntheses

• Wolfe-Quintero et al. (1998):
  – mostly ‘global’ measures (words / sentence, T-unit, clause; clauses / sentence, T-unit, etc.)
  – mostly cross-sectional
  – positive linear correlations with L2 proficiency only for T-unit length

• Ortega (2003); Norris & Ortega (2009):
  – variation and non-linearity
    • intermediate levels: more clauses / T-unit
    • advanced levels: longer clauses
Dynamic Systems Theory (DST) in L2 study

• L2 development:
  – a dynamic process, in which
  – regular growth stages are modulated by a complex variation
  – within and among individuals
  – and by continuous waxing and waning
  – of interrelated aspects of the interlanguage system (Larsen-Freeman, 2006; van Geert, 2008; Verspoor et al., 2008)
DST studies

• Dynamic relationships between **global** measures
  – among CAF dimensions
    • Larsen-Freeman (2006); Spoelman & Verspoor (2010); Gunnarsson (2012); Verspoor et al. (2012); Zhang & Lu (2013)
  – between lexical and grammatical complexity
    • Verspoor et al. (2008); Spoelman & Verspoor (2010); Vyatkina (2012)
  – between clausal and phrasal complexity
    • Spoelman & Verspoor (2010); Kormos (2011); Lu (2011)
Specific modifiers and proficiency

• Prepositions or PPs:
  – positive correlation (Connor, 1990; Ferris, 1994; Granfeldt & Nugues, 2007; Grant & Ginther, 2000; Hirschmann, in press; Hawkins & Buttery, 2010)

• Adjectives (esp. attributive):
  – positive (Granfeldt & Nugues, 2007; Grant & Ginther, 2000)

• Adverbs or adverbials:
  – positive (Ferris, 1994; Grant & Ginther, 2000; Hirschmann, in press)
  – negative (Aarts and Granger, 1998)
Specific modifiers and proficiency

- Subordinating conjunctions:
  - positive (Aarts & Granger, 1998; Granger & Rayson, 1998; Grant & Ginther, 2000; Verspoor et al., 2012)
  - negative (Ferris, 1994; Kormos, 2011)
  - none (Lu, 2011; Reid, 1992)

- Explanation for divergent findings:
  - heterogeneous category; sentence position, register, L1 effects
## Our contribution to L2 complexity research

<table>
<thead>
<tr>
<th>previous research</th>
<th>our study</th>
</tr>
</thead>
<tbody>
<tr>
<td>cross-sectional</td>
<td>longitudinal</td>
</tr>
<tr>
<td>L2 English</td>
<td>L2 German</td>
</tr>
<tr>
<td>intermediate to advanced proficiency</td>
<td>beginning proficiency</td>
</tr>
<tr>
<td>global complexity measures</td>
<td>specific complexity measures</td>
</tr>
<tr>
<td>single linguistic feature or single case studies</td>
<td>syntactic modification as a system; a learner cohort</td>
</tr>
</tbody>
</table>
The corpus:

2\textsuperscript{nd} KanDeL cohort

- longitudinal writing corpus (4 semesters)
- dense data collection intervals (3-5 weeks)
- homogenous L1 background (American English)
- all students enrolled in one program at a large public US university
- equal type and amount of exposure to L2 (mostly classroom interaction and instructional materials)
- controlled instructional conditions with documented task and learner variables
- level-appropriate writing tasks with topics and prompts reflecting the instructional content
- only rough drafts
Completed KanDeL studies

• Quasi-longitudinal data: the 1st KanDeL cohort
  – linear increase:
    • global complexity (sentence length, lexical variety)
    • range and variety of selected morphosyntactic forms
      – subordination
  – no significant increase in clause length

• Individual longitudinal data (2 learners):
  – linear increase in global complexity
  – diverging paths for specific complexity
    • learner 1: clausal; learner 2: phrasal
      • Vyatkina (2012, 2013)
## Sub-corpus composition

<table>
<thead>
<tr>
<th>Semester</th>
<th>Time (T)</th>
<th>Number of texts</th>
<th>Tokens (total)</th>
<th>Tokens (mean)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>1</td>
<td>11</td>
<td>1033</td>
<td>94</td>
<td>13.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12</td>
<td>1251</td>
<td>104</td>
<td>15.1</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>12</td>
<td>1452</td>
<td>121</td>
<td>27.2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>12</td>
<td>1114</td>
<td>93</td>
<td>20.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>883</td>
<td>88</td>
<td>5.1</td>
</tr>
<tr>
<td>2nd</td>
<td>6</td>
<td>11</td>
<td>1554</td>
<td>141</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>12</td>
<td>1638</td>
<td>137</td>
<td>33.8</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>10</td>
<td>1482</td>
<td>148</td>
<td>56.7</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>11</td>
<td>1565</td>
<td>142</td>
<td>21.7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>12</td>
<td>1669</td>
<td>139</td>
<td>30.1</td>
</tr>
<tr>
<td>3rd</td>
<td>11</td>
<td>12</td>
<td>2928</td>
<td>244</td>
<td>65.2</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>9</td>
<td>2170</td>
<td>241</td>
<td>70.0</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>12</td>
<td>1990</td>
<td>166</td>
<td>36.8</td>
</tr>
<tr>
<td>4th</td>
<td>14</td>
<td>11</td>
<td>2874</td>
<td>261</td>
<td>82.1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>12</td>
<td>2886</td>
<td>241</td>
<td>38.3</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>10</td>
<td>2111</td>
<td>211</td>
<td>61.9</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>6</td>
<td>1035</td>
<td>173</td>
<td>13.6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>185</strong></td>
<td></td>
<td><strong>29635</strong></td>
<td><strong>161</strong></td>
<td><strong>36.3</strong></td>
</tr>
</tbody>
</table>
Metadata

• Participants:
  – age: 18-22 (mean 19.5), 1 learner >30
  – gender: 7 female, 5 male
  – L2: 8 ‘true’ beginners, 4 ‘false’ beginners
  – 9 general track; 3 business track in the 4th sem.

• Tasks:
  – 17 essay prompts at 17 data collection points
  – personal narratives and descriptions; essays with explanatory elements; letters
  – mostly controlled and timed (50 min.), T11-12 and T14-16 uncontrolled and untimed
Finding modifiers in KanDeL

• Defining modifiers of different complexity via part of speech categories
  – STTS categories
  – Automatic analyses on tok level, manual corrections

• Levels of differentiation
  – Noun vs. adverbial modifiers
  – Inflectable vs. non-inflectable modifiers
  – Modifying words vs. phrases vs. clauses
Defining modifiers of different complexity via part of speech categories

<table>
<thead>
<tr>
<th></th>
<th>noun/verb</th>
<th>inflectable?</th>
<th>word/phrase/clause</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CARD</td>
<td>n</td>
<td>-</td>
<td>w</td>
</tr>
<tr>
<td>2. ADJA</td>
<td>n</td>
<td>+</td>
<td>w</td>
</tr>
<tr>
<td>3. ADJD</td>
<td>n (v)</td>
<td>+</td>
<td>w</td>
</tr>
<tr>
<td>4. ADV</td>
<td>v</td>
<td>-</td>
<td>w</td>
</tr>
<tr>
<td>5. APPR</td>
<td>v</td>
<td>-</td>
<td>p</td>
</tr>
<tr>
<td>6. PRELS</td>
<td>n</td>
<td>-</td>
<td>c</td>
</tr>
<tr>
<td>7. KOUS</td>
<td>v</td>
<td>-</td>
<td>c</td>
</tr>
</tbody>
</table>

• Many expectations; examples:
  • clausal modifiers PRELS & KOUS start at zero level, increasing
  • ADJD used in copula constructions, decreasing
  • ADJA increasing (complex inflectional system)
  • APPR always needed – no tendency?
Data analysis methods

• Calculating correlation between
  – the use of 7 individual categories
    (normalized frequencies)
  and
  – 17 points in time
    • variation at each point in time
    • individual trajectories: category use
      of each learner per text
    • group trajectory: generalized linear mixed model
      for development of whole group over time
Statistically relevant features

1. We cannot expect a linear development over the full time span
3. Two parallel problems:
   a. Specific task description at each time point.
   b. Each student has her own writing style.
Problem 1: Allowing for bended curves

This is a straight line:

\[
f(x) = 7 \cdot x + 3
\]

no bend so far
Problem 1: Allowing for bended curves

This is a parabola, a second order polynomial:

\[ f(x) = 7 \cdot x + 3 \]
\[ g(x) = f(x) - 0.16 \cdot x^2 \]

1 bend
Problem 1: Allowing for bended curves

This is a cubic curve, a third order polynomial:

\[ f(x) = 7 \cdot x + 3 \]
\[ g(x) = f(x) - 0.16 \cdot x^2 \]
\[ h(x) = g(x) + 0.015 \cdot x^3 \]

2 bends
Allowing for bending curves: Summary

• Polynomials allow for bended curves
• The higher the order of the polynomial, the more bends are possible
Model selection

• Which order of polynomial is needed?
• AIC based decision
• AIC combines two competing criteria:
  – A better model reproduces the data better: **Maximizing likelihood** of the data
  – A better model needs fewer parameters: **Minimizing the number of parameters**
Problem 2: Non-normal Response

- Only a finite number $N$ of token can be modified
- Number of modifiers is sandwiched between 0 and $N$
- Such a count is binomially distributed
- can be approximated by a normal distribution, but
  - Variance depends on the expectation value
  - Counts near the border cause serious deviations

**Solution** is the use of generalized models which model the binomial distribution (rather) directly
Problem 3: Correlated data points

- Each person has his style
- Each test subject might have a different modifier profile
- This can be described as crossed random effects.

**Solution** is mixed models as implemented in lme4:

- Bates D, Maechler M, Bolker B and Walker S (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.0-6, [http://CRAN.R-project.org/package=lme4](http://CRAN.R-project.org/package=lme4)
Remaining issues

• *Are polynomials the optimal choice?* We could use a completely nonparametric form (additive modeling)

• *Are the data really binomially distributed?* Normalization is the notorious problem with this view. Modelling text length as continuous we could switch to a Poisson distribution

• *Auto correlation of time points?*
Research question 1

1. (How) does the size and range of the syntactic modification system in the writing of 12 beginning learners of German change over two years of instructed language study?

2. What are the group developmental trends in the frequency of different modifier classes and (how) are they modulated by individual differences and observation points?
Individual learner trajectories: all modifiers
Mean frequency counts:
all modifiers

POS
- ADJA
- ADJP
- ADV
- PREP
- CARD
- SUB
- REL
RQ1: summary

• The size and range of the modification system as a whole does not change considerably over the observed period
• However, different modifiers contribute to the overall modification range to a different extent at each observation point
Research questions 2

1. (How) does the size and range of the syntactic modification system in the writing of 12 beginning learners of German change over two years of instructed language study?

2. What are the group developmental trends in the frequency of different modifier classes and (how) are they modulated by individual differences and observation points?
ADJA normalized frequencies, fitted curve, and CI
ADJA individual trajectories and mean frequencies with CI
CARD normalized frequencies, fitted curve, and CI
CARD individual trajectories
ADJP normalized frequencies, fitted curve, and CI
ADJP individual trajectories
ADV individual trajectories
PREP normalized frequencies, fitted curve, and CI
PREP individual trajectories
SUB normalized frequencies, fitted curve, and CI
SUB individual trajectories

![Graph showing individual trajectories over time with different markers and colors for each individual.](image-url)
REL normalized frequencies, fitted curve, and CI
REL individual trajectories
RQ2: Summary

• ADJA, SUB, and REL show increasing trends, CARD and ADJP decreasing trends, ADV no discernible trend, and PREP an increasing and then decreasing trend.

• Individual differences modulate group trends to a larger or smaller extent depending on the modifier type and observation point, although it is overall considerable.
Discussion

• Interplay of **stability and variability** in the system (Larsen-Freeman, 2012)

• **Stability**: syntactic modification as a global construct is a constant characteristic of L2 writing from its inception, whose size and range remain relatively stable over four semesters of language study.

• **Variability**: over time, inter-individual, and intra-individual
Clear developmental trends

• Confirm cross-sectional research results:
  – ↑ ADJA (Granfeldt & Nugues, 2007; Grant & Ginther, 2000)
  – ↓ ADJP (Hinkel, 2002)
  – ↑ SUB and REL (Byrnes & Sinicrope, 2008; Verspoor et al., 2012)

• => Benchmarks that capture development at lower levels of proficiency and between relatively dense data collection waves.
DST explanations

• Linguistic complexity and structural complexity (compositionality, embeddedness): growth in more difficult, ergo developmentally advanced (inflected, clausal) modifiers and decline in less developmentally advanced (uninflected) modifiers (= Verspoor et al., 2012; van Geert, 2008).

• Trends are not linear but curvilinear, with smoother or abrupt curves for different variables (e.g., Verspoor, 2012; Zhang & Lu, 2012).
No clear trends

• ADV: growth expected later (Hirschmann, in press)
• PREP: heterogeneous category; growth expected later with the development of the NP modification
• => ADV and PREP don’t capture development at early stages
Other effects

• topic
  – oscillations in frequency counts do not affect fitted curves => the model is working despite topic variation

• instructional focus
  – triggered or enhanced growth (ADJA, SUB, REL)

• inter-individual variation
  – learner agency, individual preferences, widening CI (cf. Verspoor et al., forthcoming)
Future research

• incorporate lexical complexity aspects
• accuracy dimension
• instructional effects and task effects
• dynamic relationships between different modifiers
• extending the observation period
• more learners – including quasi-longitudinal data and data from other cohorts
Thanks for your comments!

Acknowledgments:
• Marc Reznicek (for contributing to earlier stages of this study)
• Emily Hackmann and Michael Grünbaum (University of Kansas) for data annotation
• The German and American Fulbright Commission
• KU Institute for Digital Research in the Humanities, Digital Humanities Seed Grant
• Language Learning Small Grants Research Program

Nina Vyatkina  vyatkina@ku.edu
Hagen Hirschmann  hirschhx@hu-berlin.de
Felix Golcher  felix.golcher@hu-berlin.de
KanDeL sample search
Mixed effects results ADJA

|                | Estimate | Std. Error | z value | Pr(|z|) |
|----------------|----------|------------|---------|--------|
| (Intercept)    | -2.76    | 0.13       | -21.34  | <0.001 |
| poly(time,4)1  | 11.13    | 1.58       | 7.05    | <0.001 |
| poly(time,4)2  | -1.03    | 1.63       | -0.63   | 0.528  |
| poly(time,4)3  | -2.73    | 1.48       | -1.84   | 0.066  |
| poly(time,4)4  | -0.39    | 1.70       | -0.23   | 0.819  |

Fixed effects values for category ADJA (4\textsuperscript{th} / 4\textsuperscript{th} order)
Mixed effects results CARD

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | -3.26 | 0.22 | -14.93 | <0.001 |
| poly(time,1) | -9.56 | 2.72 | -3.52 | <0.001 |

Fixed effects values for category CARD (1\textsuperscript{st} / 1\textsuperscript{st} order)
Mixed effects results ADJP

|                 | Estimate | Std. Error | z value | Pr(>|z|) |
|-----------------|----------|------------|---------|----------|
| (Intercept)     | -3.44    | 0.08       | -44.12  | <0.001   |
| poly(time,3)1   | -2.71    | 0.92       | -2.94   | 0.003    |
| poly(time,3)2   | 2.32     | 0.99       | 2.35    | 0.019    |
| poly(time,3)3   | -1.86    | 0.87       | -2.14   | 0.033    |

Fixed effects values for category ADJP (3^{rd} / 2^{nd} order)
Mixed effects results PREP

|                | Estimate | Std. Error | z value | Pr(>|z|) |
|----------------|----------|------------|---------|----------|
| (Intercept)    | -2.73    | 0.08       | -33.68  | <0.001   |
| poly(time,2)1  | 1.30     | 1.04       | 1.25    | 0.212    |
| poly(time,2)2  | -2.39    | 1.04       | -2.31   | 0.021    |

Fixed effects values for category PREP (2\textsuperscript{nd} / 1\textsuperscript{st} order)
Mixed effects results SUB

| Estimate       | Std. Error | z value | Pr(>|z|)  |
|----------------|------------|---------|-----------|
| (Intercept)    | -4.52      | 0.33    | -13.83    | <0.001    |
| poly(time,2)1  | 26.71      | 5.02    | 5.32      | <0.001    |
| poly(time,2)2  | -8.02      | 2.82    | -2.84     | 0.004     |

Fixed effects values for category SUB (2nd / 0 order)
Mixed effects results REL

| Estimate     | Std. Error | z value | Pr(>|z|) |
|--------------|------------|---------|----------|
| (Intercept)  | -6.80      | 0.64    | -10.56   | <0.001   |
| poly(time,1) | 23.75      | 7.48    | 3.17     | 0.002    |

Fixed effects values for category REL (1st / 0 order)