Cross-linguistic effects vs. universal learning mechanisms: A case study on temporal expression

Valentin Werner (University of Bamberg)
Robert Fuchs (University of Hamburg)
Sandra Götz (Justus Liebig University Giessen)

13th Teaching and Language Corpora Conference
Cambridge, 20 July 2018
Outline

- L2-acquisition of tense-aspect (TA)
- Present Perfect (PP) in learner English(es)
- Research questions
- Data and methodology
- Results
- Discussion and conclusion

Slides and references available at educ.cam.ac.uk/events/conferences/talc2018/programme/presentations
**L2 acquisition of tense-aspect**

<table>
<thead>
<tr>
<th>L1 influence/transfer</th>
<th>Learning mechanisms/universals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Role of parallel structures</strong> in learners’ L1 (form and/or function affected)**</td>
<td><strong>Role of universal principles</strong></td>
</tr>
<tr>
<td>Presence</td>
<td>Similar rates of uptake of target structure in all learner populations (regardless of learners’ L1)</td>
</tr>
<tr>
<td>→ facilitating effect</td>
<td></td>
</tr>
<tr>
<td>→ overgeneralization</td>
<td></td>
</tr>
<tr>
<td>Absence</td>
<td></td>
</tr>
<tr>
<td>→ inhibits acquisition</td>
<td></td>
</tr>
<tr>
<td>→ undergeneralization</td>
<td></td>
</tr>
<tr>
<td><strong>Proponents</strong></td>
<td><strong>Proponents</strong></td>
</tr>
</tbody>
</table>
Present Perfect in learner English(es)

• **LCR-based studies: TA generally problematic** for learners (Granger 1999; Wulff et al. 2009; Rogatcheva 2012; Götz 2015; Howard & Leclercq 2017; Deshors 2018; Fuchs & Werner 2018)

• **Area of pervasive variation** in English in general (e.g. Werner et al. 2016)

• **PP challenging area** even for **advanced learners** of English (Klein 1995; Bardovi-Harlig 2000; Housen 2002b; Rogatcheva 2014)
Present Perfect in learner English(es)

• Case study of L1 German EFL learners
  (Fuchs et al. 2016; see also Davydova 2011; Kämmerer 2012; Rogatcheva 2014)
  – Morphological (not functional/semantic!) similarity:
    \text{HAVE/HABEN} + \text{past participle}
    
    \textit{(He has bought a car vs. Er hat ein Auto gekauft)}
  – BUT \text{underuse} of PP in English \textit{(pace L1-influence position)}

• Situation for learner varieties where L1 \text{doesn’t} have a
  (morphologically) PP-like structure, but a perfective particle
  – PP as a (generally) \textbf{challenging} structure?
  – \textbf{Mediating influence} of L1?
Present Perfect in learner English(es): Research questions

1. **The PP as challenging structure:**
   Which rates of uptake do learners show?
   How and where do they differ from native speaker usage?

2. **SLA principles:** Is the acquisition of TA and the difference to native speaker usage influenced by (i) cross-linguistic effects, (ii) guided by universal principles (irrespective of the learners’ L1), or rather by an interaction of (i) and (ii)?

3. **Linguistic principles:** Can the use of PP/SP be predicted by cognitive-linguistic variables, such as priming effects, *Aktionsart* (lexical aspect) of the verb, etc.?
Data and methodology

Research design

• “SLA-driven” (Myles 2015) study (testing transfer)

• **Contrastive Interlanguage Analysis** (Granger 1996, 2015)
  L1 German learners (PP-like form in L1)
  vs.
  L1 Cantonese learners (no PP-like form in L1; Yue 2003)
  vs.
  Native speakers (US/GB)

• **Mode**: spoken vs. written data
## Data and methodology

<table>
<thead>
<tr>
<th></th>
<th>Learner English I</th>
<th>Learner English II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(German Learner English)</td>
<td>(Chinese Learner English)*</td>
</tr>
<tr>
<td>c. 400,700 words total</td>
<td></td>
<td>c. 542,500 words total</td>
</tr>
<tr>
<td><strong>GICLE</strong></td>
<td>Learner writing of university</td>
<td>Learner writing of university</td>
</tr>
<tr>
<td></td>
<td>students</td>
<td>students</td>
</tr>
<tr>
<td><strong>LINDSEI-GE</strong></td>
<td>Learner speech of university</td>
<td>Learner speech of university</td>
</tr>
<tr>
<td></td>
<td>students</td>
<td>students</td>
</tr>
<tr>
<td>c. 234,000</td>
<td>c. 86,100</td>
<td>c. 412,900</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. 65,000</td>
</tr>
</tbody>
</table>

*NB: data comprises speakers of Southern Chinese dialects (mainly Cantonese) only, other data excluded*
Data and methodology

<table>
<thead>
<tr>
<th>Native English (British and American English)</th>
<th>c. 433,700 words total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOCNESS-GB</td>
<td>LOCNESS-US</td>
</tr>
<tr>
<td>Native writing of British school students (A-level)</td>
<td>Native writing of British university students</td>
</tr>
<tr>
<td>c. 55,049</td>
<td>c. 92,254</td>
</tr>
</tbody>
</table>
Methodology

Corpus processing
• All corpora tagged with CLAWS C7 tagset (Garside & Smith 1997)
• *AntConc* (Anthony 2014) for corpus queries through regular expressions based on POS
• 3,754 PP tokens, 19,376 SP tokens
• Only full VPs (subject + finite verb): Excluded repetitions, self-corrections, irrealis (e.g. *if I were you*), possessive *got*, unclear cases
Methodology

Annotation
Annotated all PP tokens and subset of SP tokens (500 per subcorpus) for the following variables

1. **Verb semantics**: Activity (e.g. *look*), aspectual (e.g. *start*), causative (e.g. *make*), communication (e.g. *answer*), existence (e.g. *live*), mental (e.g. *believe*), occurrence (e.g. *happen*) (following Deshors 2017)

2. **Aktionsart**: Accomplishment (e.g. *run a mile*), achievement (e.g. *discover*), activity (e.g. *run*), stative (e.g. *believe*) (following Katz 2003 and Vendler 1967)

3. **Time adverbial**:
   - Definite (e.g. *last year*), indefinite (e.g. *always, ever*), no time adverbial

4. **Mode of the verb**: Active, passive

5. **Progressive**: Yes, no

6. **Verb class**: regular, irregular

7. **Negation**: Yes, no

8. **Preceding tense/priming**:
   - Simple past, present perfect, present, modal, other
Methodology

Statistical analysis

• MuPDAR
  (Multifactorial Prediction and Deviation Analysis with Regressions; Gries & Deshors 2014; Gries & Bernaisch 2016)

• Assesses resemblances and differences between native speakers (“target”) and learners in linguistic choices (here: choice between PP and SP)

• As well as between distinct groups of learners (L1 Cantonese & L1 German → transfer?)
Methodology

Statistical analysis: MuPDAR

1. Regression $R_1$: Native data
   Statistical model of the factors that influence the linguistic choice in native usage

2. $R_1$ applied to learner data
   Predict for all instances:
   Would a native speaker have made the same choice under similar circumstances?

3. Regression $R_2$:
   Regression on residuals from step 2
   How do the learners’ choices differ from native usage and which linguistic variables can account for this difference?
   Here: Classification and Regression Trees (CART) and random forests
Methodology

Statistical analysis: MuPDAR – Example

Ex. 1: Learner chooses SP

-> Native speaker would have chosen SP
-> Factor time adverbials responsible
("I worked yesterday")
Methodology

Statistical analysis: MuPDAR – Example

Ex. 1: Learner chooses SP
-> Native speaker would have chosen SP
-> Factor time adverbials responsible
   (“I worked yesterday”)

Ex. 2: Learner chooses PP
-> Native speaker would have chosen SP
-> Factor time adverbials responsible
   (“I have worked yesterday”)

Results I: Variable importance

Whether learners deviate from native usage is explained by the following variables:
Results II: Specific contexts of usage

Low likelihood of errors

• SP, primed/preceded by SP, time adverbial definite/none

  *She drove to a house, jumped out, put the letter in the letter box and [...] (LINDSEI GEAU 1040)*

• PP, primed/preceded by P/PP/mod/other, verb semantics occurrence/causative/active/aspectual

  *once you realise that they have done so much to em help you (LINDSEI CH 027)*

• Account for +50% of the data
Results II: Specific contexts of usage

High(er) likelihood of errors

• PP, primed/preceded by SP, time adverbial definite/none (error gravity/deviation slightly less problematic with activity verbs than other verb semantics)

  *it was in November when we’ve been there* (GICLE AU3026)

• SP, primed/preceded by P/PP, verb semantics occurrence/causative/activity/aspectual

  *very unfortunately I seldom travel yeah the the farthest place I traveled is is er Guangzhou* (LINDSEI CH012)
1. **The PP as challenging structure**: Which rates of uptake do learners show? How and where do they differ from native speaker usage?
Discussion

1. The PP as challenging structure: Which rates of uptake do learners show? How and where do they differ from native speaker usage?
   • High degree of native-like uptake
   • Specific contexts of non-native-like usage identified (e.g. when the context included an indefinite or no time adverbial, choosing the SP yields a non-native-like pattern, with a higher error gravity in spoken than in written language)
Discussion

2. **SLA principles**: Is the acquisition of TA and the difference to native speaker usage influenced by (i) cross-linguistic effects, (ii) guided by universal principles (irrespective of the learners’ L1), or rather by an interaction of (i) and (ii)?
Discussion

2. **SLA principles:** Is the acquisition of TA and the difference to native speaker usage influenced by (i) cross-linguistic effects, (ii) guided by universal principles (irrespective of the learners’ L1), or rather by an interaction of (i) and (ii)?

- **Cognitive/linguistic variables** explain speakers’ use of a PP vs. an SP construction: priming effects, the semantics of the word class, time adverbial, mode

- **L1 not selected** as a factor → no cross-linguistic differences in linguistically conditioned error likelihood (L1 Cant. more likely to make errors, but no differences in linguistic conditioning)
Discussion

3. Linguistic principles: Can the use of PP/SP be predicted by cognitive-linguistic variables, such as priming effects, Aktionsart (lexical aspect) of the verb, etc.?
3. **Linguistic principles:** Can the use of PP/SP be predicted by cognitive-linguistic variables, such as priming effects, *Aktionsart* (lexical aspect) of the verb, etc.?

Yes.
Conclusion

• Fine-grained analysis of linguistic conditioning of native-like and non-native like usage of PP/SP

• No cross-linguistic differences in the conditioning, only error likelihood, despite cross-linguistic differences in similar structures

• Supports universal learning mechanisms as explanation
Thank you very much!

Looking forward to all kinds of questions and comments.

*******************************************************************************

Thanks to Kathrin Kircili for her invaluable help with the annotation of the verb forms and to Fabian Vetter for tagging the corpus data!
Cross-linguistic effects vs. universal learning mechanisms:
A case study on temporal expression

Corpora used
ICCI (Tono & Diez-Bedmar 2014)
ICLE (Granger, Dagneaux, Meunier & Paquot 2009)
LINDSEI (Gilquin, De Cock & Granger 2010)
LOCNEC (De Cock 2004)

References


