THE GERMAN LEARNER MIDDLE FIELD

HOW TO ACQUIRE HIDDEN FACTORS CORPUS STUDY ON THE FALKO ADVANCED LEARNER CORPUS

Marc Reznicek
Humboldt-Universität zu Berlin

AELFE 2011 – Universidad Politècnica de Valencia
5.-7.9.2011
Overview

• Acquiring linguistic variation
• The German middle field
• Variation in the German middle field
• Modeling
• Falko
• Annotation
• Analysis
• Results
• Outlook
Acquiring linguistic variation

- Long tradition of syntax acquisition research (see Ellis 2009)
Acquiring linguistic variation

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005)
Acquiring linguistic variation

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005)
- Only few studies on acquisition of variation in syntactic patterns
Acquiring linguistic variation

- Long tradition of syntax acquisition research (see Ellis 2009)
- Focus mainly on acquisition of word order rules and acquisition stages (e.g. Pienemann 2005)
- Only few studies on acquisition of variation in syntactic patterns
- Research question:
  How do second language learners acquire the competence for using those competing structures?
German topological field model

- Topological field model for German (Drach 1937, Höhle 1986, Pasch et al. 2003)

<table>
<thead>
<tr>
<th>pref</th>
<th>lsb</th>
<th>MF</th>
<th>rsb</th>
<th>post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Der Feminismus</td>
<td>hat</td>
<td>den Frauen schon immer</td>
<td>geschadet</td>
<td>durch seine Radikalität</td>
</tr>
<tr>
<td><em>The feminism-NOM</em></td>
<td><em>has</em></td>
<td><em>the women-ACC</em></td>
<td><em>damaged</em></td>
<td><em>with its radicality</em></td>
</tr>
</tbody>
</table>

lsb: left sentence bracket
rsb: right sentence bracket
German topological field model

- Topological field model for German (Drach 1937, Höhle 1986, Pasch et al. 2003)
- Verb-Second Rule (V2)

<table>
<thead>
<tr>
<th>pref</th>
<th>lsb</th>
<th>MF</th>
<th>rsb</th>
<th>post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Der Feminismus</td>
<td>hat</td>
<td>den Frauen schon immer</td>
<td>geschadet</td>
<td>durch seine Radikalität</td>
</tr>
<tr>
<td>The feminism-NOM</td>
<td>has</td>
<td>the women-ACC</td>
<td>damaged</td>
<td>with its radicality</td>
</tr>
</tbody>
</table>

lsb: left sentence bracket
rsb: right sentence bracket
## Variation in the German middle field

- **scrambling:**

Constituents in the middle field allow a variety of competing word orders (Haider/Rosengreen 2003)

<table>
<thead>
<tr>
<th>dass</th>
<th>[<strong>diese Ansicht</strong>]AKK [in Zukunft] [<strong>viel mehr Menschen</strong>]NOM</th>
<th>zu teilen lernen</th>
</tr>
</thead>
<tbody>
<tr>
<td>that</td>
<td>[<strong>those opinions</strong>]ACC [in the future] [a lot more people] NOM</td>
<td>to share learn</td>
</tr>
</tbody>
</table>

(dew07_2007_09_v2.1)

<table>
<thead>
<tr>
<th>dass</th>
<th>[<strong>viel mehr Menschen</strong>]NOM [in Zukunft] [<strong>diese Ansicht</strong>]AKK</th>
<th>zu teilen lernen</th>
</tr>
</thead>
<tbody>
<tr>
<td>dass</td>
<td>[in Zukunft] [<strong>viel mehr Menschen</strong>]NOM [<strong>diese Ansicht</strong>]AKK</td>
<td>zu teilen lernen</td>
</tr>
</tbody>
</table>

AELFE 2011  The German Learner Middle Field  Marc Reznicek  Humboldt-Universität zu Berlin
Factors in middle field word order

- Word order is not strictly rule based
Factors in middle field word order

- Word order is not strictly rule based
- A variety of **influencing factors** for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)
Factors in middle field word order

- Word order is not strictly rule based
- A variety of influencing factors for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)

- grammatical function
  subject., dir. object., ind. object
- case
  nominative, accusative, dative
- part-of-speech
  personal pronoun, full noun, reflexive
- weight
  amount of word, amount of syllables
Factors in middle field word order

- Word order is not strictly rule based
- A variety of **influencing factors** for word orders have been discussed (e.g. Siewierska 1997, Uszkoreit 1987)

- **grammatical function**
  - subject., dir. object., ind. object
- **case**
  - nominative, accusative, dativ
- **part-of-speech**
  - personal pronoun, full noun, reflexive
- **weight**
  - amount of word, amount of syllables
- **phrase type**
  - noun phrase, prepositional phrase, clause
- **semantic role**
  - agent, patient, recipient
- **information status**
  - given, new
- **agentivity**
  - person, institution, animal, materia
Modeling competing factors

- Most factors have been looked at one at a time
  (see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
Modeling competing factors

- Most factors have been looked at one at a time
  (see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors
Modeling competing factors

- Most factors have been looked at one at a time
  (see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors
- Possibility I: Hierarchies
  - Optimality theory (Uzkoreit 1987)
Modeling competing factors

- Most factors have been looked at one at a time
  (see Kurz 2000, Heylen et al. 2005, Bader/Häusler 2010)
- For modeling of simultaneous influence of competing factors
- Possibility I: Hierarchies
  - Optimality theory (Uzkoreit 1987)
- Possibility II: Relative factor strength analysis
L1 results for news paper articles

- Grammatical function has a strong effect
  - 96% SB-OB   4% OB-SB

(Bader & Häusler 2010)
Grammatical function has a strong effect
- 96% SB-OB  4% OB-SB

Case influences word order in NN-NN combinations

SB – ACC_{OBJ} (99%) > SB – DAT_{OBJ} (75%)
L1 results for news paper articles

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB  4% OB-SB

- Case influences word order in NN-NN combinations
  - SB – ACC<sub>OBJ</sub> (99%) > SB – DAT<sub>OBJ</sub> (75%)

- Part-of-Speech has a strong effect
  - pronouns > full nouns
L1 results for news paper articles

(Bader & Häusler 2010)

- Grammatical function has a strong effect
  - 96% SB-OB  4% OB-SB
- Case influences word order in NN-NN combinations
  - SB – ACC<sub>OBJ</sub> (99%) > SB – DAT<sub>OBJ</sub> (75%)
- Part-of-Speech has a strong effect
  - pronouns > full nouns
- Constituent-weight has no effect
Research Question:

Do second language learner texts show a difference in effect strength for those factors than native speaker texts?
Research Question:

Do second language learner texts show a difference in effect strenght for those factors than native speaker texts?

- Contrastive Interlanguage analysis CIA (Granger 2008)
  - Assumption
    - learner language is systematic
    - variation in the group
    - transfer & generell language acquisition processes
Data: Falko learner corpus of German

- advanced learners of German B1+
- essays and summaries
- cross-sectional & longitudinal data
- ~260,000 tokens, growing
- automatically annotated POS, lemma
  (Treetagger, Schmid 1994)
- dependency parsed (NEW)  (Bohnet 2010)

Lüdeling et al. 2008
Data: Falko learner corpus of German

- advanced learners of German B1+
- essays and summaries
- cross-sectional & longitudinal data
- ~260,000 tokens, growing
- automatically annotated POS, lemma
  (Treetagger, Schmid 1994)
- dependency parsed (NEW)  (Bohnet 2010)

Subset used
- 94 texts learners of German  (25 L1s)
- 94 text  German control group

http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko/standardseite/
Non-canonical syntactic structures in learner texts (LT) make a description with standard grammars impossible.

*LT: Aber in die meisten Fällen das **ist** nicht der Fall.*

(FalkoEssayL2v2.0:fk006_2006_08)

*But unfortunately such percentages define the value of universities.*
Therefore a minimal grammatical correction (TH1) is explicitly included into the corpus (Reznicek et al. 2009)

TH1: Aber in den meisten Fällen ist das nicht der Fall.

LT: Aber in die meisten Fällen das ist nicht der Fall.

But in the-FEM most cases-MASC that is not the case.
To conserve the original word order, dependencies are mapped back on original sites.

\textbf{TH0}: Aber in \textit{den} meisten Fällen \textit{das} \textit{ist} nicht der Fall.

\textbf{TH1}: Aber in \textit{den} meisten Fällen \textit{ist} \textit{das} nicht der Fall.

\textbf{LT}: Aber in \textit{die} meisten Fällen \textit{das} \textit{ist} nicht der Fall.

But in the-FEM most cases-MASC that is not the case.
**THO:** Aber in den meisten Fällen das ist nicht der Fall.

But in the-FEM most cases-MASC that is not the case.
Data : Target hypotheses

Each dependency is automatically labeled with the sentence function.

\[ \text{THO: } \textcolor{red}{\text{das}} \textcolor{black}{\text{ ist}} \textcolor{black}{\text{nicht der Fall.}} \]

...that is not the case.
Annotation : middle fields

- In all utterances the middle fields have been manually annotated.
Annotation: middle fields

- In all utterances the middle fields have been manually annotated.
- For each middle field, following information has been extracted.
Annotation : middle fields

- In all utterances the middle fields have been manually annotated.
- For each middle field following information has been extracted.
- Only for verb arguments
  1) clause type (main clause, subordinate clause)
  2) verb argument order (obj-sub, sub-obj)
  3) part-of-speech (noun, pron, prf, prep)
  4) case (nom, acc, dat)
  5) length of constituent in tokens
  6) length of constituents in sillables
method: linear mixed effect model

linear mixed effect model to calculate the effect strength of different factors:

\[ z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_{k+1} \]

\rightarrow \text{probabilities for OB-SB-order with subject as full noun random effects: verb, text}
results I: $\chi^2$

- Learners use significantly less object-subject middle fields in subordinate clauses
results I: $\chi^2$

- Interestingly this is not the case in main clauses
results II: effects & interactions

We look for interactions of L2 with other factors
Only interaction: language & part-of-speech when reflexive pronoun
results II: effects & interactions

L1

main clause

subordinate clause

L2

AELFE 2011 The German Learner Middle Field Marc Reznicek Humboldt-Universität zu Berlin
The learners in this study have shown less variation in the use of SB-OB-type subordinate clauses. This seems to mainly come from a significant bias of SB-OB-type clauses for reflexive pronouns.
discussion

- The learners in this study have shown less variation in the use of SB-OB-type subordinate clauses.
  - This seems to mainly come from a significant bias of SB-OB-type clauses for reflexive pronouns.

- **NO** effect found for case, weight.
  - **case**: Too few datives in the data.
  - **weight**: cognitive load ➔ language independent
summary and outlook

- Advanced learners of German show different patterns of variation linked to the verb argument order in the German middle field.
- This seems to be due to a non-native-like weight of the factors 'sentence function' and 'part-of-speech' as influence of argument order.
summary and outlook

- Advanced learners of German show different patterns of variation linked to the verb argument order in the German middle field.
- This seems to be due to a non-native like weight of the factors 'sentence function' and 'part-of-speech' as influence of argument order.

Next step:
- more semantic and pragmatic factors
Thanks to

Felix Golcher
Berlin corpus linguistics team
bibliography

- **Bates, Douglas; Maechler, Martin; Bolker, Ben (2011)**: lme4: Linear mixed-effects models using S4 classes. URL: http://CRAN.R-project.org/package=lme4
- **Drach, Erich (1937)**: Grundgedanken der deutschen Satzlehre. Frankfurt am Main: Diesterweg.
- **Lüdeling, Anke; Doolittle, Seanna; Hirschmann, Hagen; Schmidt, Karin; Walter, Maik (2008)**: Das Lernerkorpus Falko. In: Deutsch als Fremdsprache 45 (2), p.67–73.
bibliography


all sources checked on 09-05-2011.