Underuse of Syntactic Categories in Falko - A Case Study on Modification

Hagen Hirschmann
Anke Lüdeling
Ines Rehbein
Marc Reznicek
Amir Zeldes

LEARNER CORPUS RESEARCH 2011
LOUVAIN-LA-NEUVE
research questions & approach

• how can syntactic analyses of L2 learner data help in understanding interlanguage/acquisition processes?
• what is the relationship between lexical elements and syntactic classes?
  ➢ phenomenon: modification
  ➢ data: dependency-parsed corpus of advanced L2 learners of German
  ➢ CIA study (underuse statistics)
• freely available annotated learner corpus of German as a foreign language
• advanced learners (tutored acquisition)
• written language / controlled, unaided writing
• several text types (sub-corpora); here essays (ca. 130000 tokens)
• comparable native speaker corpora (ca. 70000 tokens)
• meta-data for each learner (bibliographic data, linguistic history, c-test score)
• Lüdeling et al. (2008), Reznicek et al. (2010), http://www.linguistik.hu-berlin.de/institut/professuren/-korpuslinguistik/forschung/falko/standardseite
annotations in Falko

• standoff format (token annotation, span annotation, graphs, pointers etc.), annotation layers can be freely added (Lüdeling et al. 2005)

• learner utterance
  – pos & lemma (automatic, manual correction)
    (TreeTagger, Schmid 1994)
  ➢ **target hypotheses** (manual, as many as necessary)
    – pos & lemma
    – error annotation (automatic)
    – parses (dependencies; automatic, manual correction)
    – manual error annotation of some phenomena
    – ...

4
annotation of learner data: conceptual issues

• annotation of learner data is highly problematic
  – data is not systematic according to L1 grammar (especially if there are different L1s)
  – difficult for automatic tools (taggers, parsers)
  – for error analysis and contrastive interlanguage analysis: data has to be interpreted

conceptual problems: pos

- word forms in L2 data sometimes correspond to different pos (Diaz-Negrillo et al. 2010)

<table>
<thead>
<tr>
<th>Most</th>
<th>important</th>
<th>of</th>
<th>all</th>
<th>was</th>
<th>the</th>
<th>conscious</th>
<th>that</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBS</td>
<td>JJ</td>
<td>IN</td>
<td>DT</td>
<td>VBD</td>
<td>DT</td>
<td>JJ</td>
<td>IN/that</td>
</tr>
<tr>
<td>most</td>
<td>important</td>
<td>of</td>
<td>all</td>
<td>be</td>
<td>the</td>
<td>conscious</td>
<td>that</td>
</tr>
</tbody>
</table>

- every assignment of a pos is an interpretation (conscious/NN?JJ → consciousness/NN)
conceptual problems: syntax

- no possible/useful parse of this structure
- utterance must be transformed into a canonical structure (Hirschmann et al. 2007)

- target hypothesis
parsing approach:
target hypotheses

- note: conflicting `th` may be formulated:
annotation of learner data: target hypothesis in Falko

• th1: sentence-based, very close to original text, mainly 'genuine' grammatical errors
• th2: text-based, also stylistic errors
• the differences between a target hypothesis and the original data is automatically annotated with edit tags (change, insert, replace etc.)

• (Lüdeling 2011, Reznicek et al. submitted)
target hypotheses …

• are just as necessary for L1 data, btw
research question

• we want to find **structural** features/problems in German L2 interlanguage

• structural problems are those problems that
  – occur independent of the learners' L1
  – and are therefore attributed to the structure of the target grammar
underuse

• L2 distributions are compared to L1 distributions
• overuse, underuse are defined as (statistically significant) differences between the varieties
• a category can be underused in L2 because
  – the learners do not know it
  – the learners do know it but (unconsciously) avoid it
underuse

- L2 distributions are compared to L1 distributions
- overuse, underuse are defined as (statistically significant) differences between the varieties
- a category can be underused in L2 because
  - the learners do not know it
  - the learners do know it but (unconsciously) avoid it
    → a diagnostics for detecting structural acquisition problems
visualization of overuse and underuse

• underuse: cold colours
• overuse: warm colours
• intensity of colour signals strength of overuse/underuse

- Excel add in by Amir Zeldes available at http://korpling.german.hu-berlin.de/~amir/uoaddin.htm
visualization of overuse and underuse: lexical categories

<table>
<thead>
<tr>
<th>lemma</th>
<th>tot_norm</th>
<th>de</th>
<th>da</th>
<th>en</th>
<th>fr</th>
<th>pl</th>
<th>ru</th>
</tr>
</thead>
<tbody>
<tr>
<td>in</td>
<td>0.013188</td>
<td>0.012261</td>
<td>0.014041</td>
<td>0.014247</td>
<td>0.015272</td>
<td>0.012135</td>
<td>0.009534</td>
</tr>
<tr>
<td>es</td>
<td>0.010897</td>
<td>0.011945</td>
<td>0.010900</td>
<td>0.011379</td>
<td>0.013347</td>
<td>0.008163</td>
<td>0.12385</td>
</tr>
<tr>
<td>sie</td>
<td>0.010618</td>
<td>0.008193</td>
<td>0.010643</td>
<td>0.008835</td>
<td>0.010909</td>
<td>0.006067</td>
<td>0.005613</td>
</tr>
<tr>
<td>man</td>
<td>0.010164</td>
<td>0.007900</td>
<td>0.012438</td>
<td>0.008742</td>
<td>0.009754</td>
<td>0.006950</td>
<td>0.007306</td>
</tr>
<tr>
<td>dass</td>
<td>0.009522</td>
<td>0.007404</td>
<td>0.012823</td>
<td>0.008789</td>
<td>0.009625</td>
<td>0.008880</td>
<td>0.009890</td>
</tr>
<tr>
<td>von</td>
<td>0.007982</td>
<td>0.007122</td>
<td>0.007309</td>
<td>0.006846</td>
<td>0.007315</td>
<td>0.010259</td>
<td>0.007930</td>
</tr>
<tr>
<td>auch</td>
<td>0.007028</td>
<td>0.008362</td>
<td>0.008527</td>
<td>0.005828</td>
<td>0.005775</td>
<td>0.005461</td>
<td>0.004455</td>
</tr>
<tr>
<td>für</td>
<td>0.006683</td>
<td>0.007201</td>
<td>0.006091</td>
<td>0.007216</td>
<td>0.006802</td>
<td>0.005736</td>
<td>0.004338</td>
</tr>
<tr>
<td>sind</td>
<td>0.006465</td>
<td>0.004271</td>
<td>0.008976</td>
<td>0.007308</td>
<td>0.006930</td>
<td>0.004964</td>
<td>0.005346</td>
</tr>
<tr>
<td>sich</td>
<td>0.006309</td>
<td>0.011697</td>
<td>0.006283</td>
<td>0.006291</td>
<td>0.006930</td>
<td>0.007170</td>
<td>0.005435</td>
</tr>
<tr>
<td>ich</td>
<td>0.006262</td>
<td>0.003877</td>
<td>0.013272</td>
<td>0.005366</td>
<td>0.003465</td>
<td>0.001434</td>
<td>0.001426</td>
</tr>
<tr>
<td>aber</td>
<td>0.006048</td>
<td>0.003347</td>
<td>0.007309</td>
<td>0.006245</td>
<td>0.007315</td>
<td>0.003365</td>
<td>0.003831</td>
</tr>
</tbody>
</table>

*sich* (reflexive pronoun) is underused in all L1 groups
visualization of overuse and underuse: bigrams of pos-categories

<table>
<thead>
<tr>
<th>bigram</th>
<th>tot_norm</th>
<th>de</th>
<th>da</th>
<th>en</th>
<th>fr</th>
<th>pl</th>
<th>ru</th>
</tr>
</thead>
<tbody>
<tr>
<td>$.-PPER</td>
<td>0.042384</td>
<td>0.005297</td>
<td>0.009748</td>
<td>0.007963</td>
<td>0.006166</td>
<td>0.005801</td>
<td>0.007409</td>
</tr>
<tr>
<td>VVFIN-$,$</td>
<td>0.042131</td>
<td>0.006457</td>
<td>0.00776</td>
<td>0.006343</td>
<td>0.006937</td>
<td>0.006243</td>
<td>0.008391</td>
</tr>
<tr>
<td>PPOSAT-NN</td>
<td>0.041739</td>
<td>0.008058</td>
<td>0.007247</td>
<td>0.007269</td>
<td>0.007066</td>
<td>0.006298</td>
<td>0.005802</td>
</tr>
<tr>
<td>ADV-ADV</td>
<td>0.041604</td>
<td>0.012858</td>
<td>0.010518</td>
<td>0.006111</td>
<td>0.006166</td>
<td>0.003094</td>
<td>0.002856</td>
</tr>
<tr>
<td>ADV-APP'R</td>
<td>0.039742</td>
<td>0.009117</td>
<td>0.008016</td>
<td>0.005324</td>
<td>0.007837</td>
<td>0.004807</td>
<td>0.004642</td>
</tr>
<tr>
<td>PDAT-NN</td>
<td>0.03956</td>
<td>0.005409</td>
<td>0.004233</td>
<td>0.005509</td>
<td>0.007837</td>
<td>0.007735</td>
<td>0.008837</td>
</tr>
<tr>
<td>ADV-ART</td>
<td>0.037125</td>
<td>0.007629</td>
<td>0.006349</td>
<td>0.006898</td>
<td>0.005653</td>
<td>0.006133</td>
<td>0.004463</td>
</tr>
</tbody>
</table>

adverb chains are underused in all L1 groups
modification

• corpus-based studies of adverbs in GFL
  – typically based on lexical items and (rarely) word classes (form-based)
  – typically for one language pair (Möllering 2004, Vyatkina 2007 etc.)
• ADV underuse points to a more general phenomenon: modification
modification

- are the effects form-based or function-based?
  - are all adverbs underused?
  - are certain adverbs (forms) underused?
  - are certain adverbs (forms) underused in certain functions?
  - are certain adverbial functions underused?
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)
modification

- are the effects form-based or function-based?
  - are all adverbs underused?
    - no; auch, noch etc. overused
  - are certain adverbs (forms) underused?
    - yes
  - are certain adverbial functions underused?
  - are certain adverbs (forms) underused in certain functions?
  - is modification generally underused?
    (or do learners make up for the underuse of adverbs by other means of modification?)
underuse of adverbs: function

• pos tag ADV is not fine-grained enough
  ➢ better classification, different functions
    – classes show different distributions
    – only some of these classes are underused by the learners

• Hirschmann (2011, in preparation)
strength of underuse of different syntactic ADV classes

PTK: particles (sehr gut - very good)

ADVV: modal adverbs (Bald schneit es – Soon it will snow)

ADVS: sentence adverbs (Bestimmt schneit es bald – Certainly, it will snow soon)

PTKM: modal particles (Es schneit wohl gerade – It is apparently snowing now)
underuse of adverbs: function

- underuse differences between different adverbial functions
- but classification still word based
- compensation strategies?
  - necessity to code syntactic functions independent of filler category
Falko – syntactic annotation

- target hypothesis of Falko L1 and L2 corpora
- manually corrected pos tags
- semi-automatic sentence segmentation
- dependency parser by Bernd Bohnet (2010; Syntactic Analyser)
- training data: TiGer dependency bank (derived from ~50000 trees of the TiGer treebank)
- result: very accurate dependency parses with syntactic functions
syntax schema (very briefly)

• every word is connected with its dependent(s)
• arrows point to hierarchically lower dependent
• each arrow (dependency) has a function label
searching for modification in Falko

• different aspects of the problem
  – is the syntactic function ‘modification’ underused?
  – what is the target of the modification?
  – what are the categories used for modification?

This is true especially for children
polyfunctional lexemes: so

So geht es aber leider nicht immer. It does not always work "so".

Dieses Fach ist so dynamisch.
This subject is so dynamic.

Das ist keine so einfache Frage, die This question is not "so" simple (, which .)
Modification

- Are the effects form-based or function-based?
  - Are all adverbs underused?
    - No; *auch, noch* etc. overused
  - Are certain adverbs (forms) underused?
    - Yes
  - Are certain adverbs (forms) underused in certain functions?
    - Yes
  - Are certain adverbial functions underused?
  - Is modification generally underused?
    - (Or do learners make up for the underuse of adverbs by other means of modification?)
overuse / underuse of syntactic functions

<table>
<thead>
<tr>
<th>label</th>
<th>de</th>
<th>da</th>
<th>en</th>
<th>fr</th>
<th>ru</th>
<th>usb</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK</td>
<td>0.26407</td>
<td>0.278546</td>
<td>0.284881</td>
<td>0.303271</td>
<td>0.29552</td>
<td>0.295136</td>
</tr>
<tr>
<td>HD</td>
<td>0.156192</td>
<td>0.155622</td>
<td>0.157178</td>
<td>0.154275</td>
<td>0.15809</td>
<td>0.156483</td>
</tr>
<tr>
<td>MO</td>
<td>0.141968</td>
<td>0.12789</td>
<td>0.113704</td>
<td>0.110112</td>
<td>0.112513</td>
<td>0.108707</td>
</tr>
<tr>
<td>SB</td>
<td>0.07398</td>
<td>0.078506</td>
<td>0.077099</td>
<td>0.075093</td>
<td>0.078852</td>
<td>0.085512</td>
</tr>
<tr>
<td>CJ</td>
<td>0.059604</td>
<td>0.053397</td>
<td>0.056411</td>
<td>0.050632</td>
<td>0.059274</td>
<td>0.072183</td>
</tr>
<tr>
<td>AC</td>
<td>0.057051</td>
<td>0.059317</td>
<td>0.057215</td>
<td>0.054796</td>
<td>0.054012</td>
<td>0.04916</td>
</tr>
<tr>
<td>OC</td>
<td>0.050335</td>
<td>0.053039</td>
<td>0.050008</td>
<td>0.049888</td>
<td>0.047125</td>
<td>0.040679</td>
</tr>
<tr>
<td>OA</td>
<td>0.044213</td>
<td>0.042352</td>
<td>0.044097</td>
<td>0.043643</td>
<td>0.046119</td>
<td>0.046218</td>
</tr>
<tr>
<td>CD</td>
<td>0.026549</td>
<td>0.024632</td>
<td>0.025639</td>
<td>0.022156</td>
<td>0.024917</td>
<td>0.030466</td>
</tr>
<tr>
<td>CP</td>
<td>0.017653</td>
<td>0.021732</td>
<td>0.020325</td>
<td>0.018141</td>
<td>0.017256</td>
<td>0.014887</td>
</tr>
<tr>
<td>PD</td>
<td>0.014435</td>
<td>0.014462</td>
<td>0.015943</td>
<td>0.015019</td>
<td>0.016947</td>
<td>0.018002</td>
</tr>
<tr>
<td>NG</td>
<td>0.011065</td>
<td>0.011561</td>
<td>0.010914</td>
<td>0.00974</td>
<td>0.00975</td>
<td>0.011252</td>
</tr>
<tr>
<td>MNR</td>
<td>0.010995</td>
<td>0.013707</td>
<td>0.013429</td>
<td>0.013338</td>
<td>0.010679</td>
<td>0.009521</td>
</tr>
<tr>
<td>RC</td>
<td>0.010051</td>
<td>0.008979</td>
<td>0.009385</td>
<td>0.011375</td>
<td>0.006268</td>
<td>0.005366</td>
</tr>
</tbody>
</table>
overuse / underuse of syntactic functions – significant results

<table>
<thead>
<tr>
<th>label</th>
<th>de</th>
<th>da</th>
<th>en</th>
<th>fr</th>
<th>ru</th>
<th>usb</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK</td>
<td>0.264067</td>
<td>0.278546</td>
<td>0.284881</td>
<td>0.303271</td>
<td>0.29552</td>
<td>0.295136</td>
</tr>
<tr>
<td>HD</td>
<td>0.156192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MO</td>
<td>0.141968</td>
<td>0.12789</td>
<td>0.113704</td>
<td>0.110112</td>
<td>0.112513</td>
<td>0.108707</td>
</tr>
<tr>
<td>SB</td>
<td>0.073968</td>
<td>0.078506</td>
<td></td>
<td></td>
<td>0.078852</td>
<td>0.085512</td>
</tr>
<tr>
<td>CJ</td>
<td>0.059604</td>
<td>0.053397</td>
<td>0.056411</td>
<td>0.050632</td>
<td></td>
<td>0.072183</td>
</tr>
<tr>
<td>AC</td>
<td>0.057051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.04916</td>
</tr>
<tr>
<td>OC</td>
<td>0.050335</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.040679</td>
</tr>
<tr>
<td>OA</td>
<td>0.044213</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>0.026549</td>
<td></td>
<td></td>
<td>0.022156</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>0.017653</td>
<td>0.021732</td>
<td>0.020325</td>
<td></td>
<td></td>
<td>0.016947</td>
</tr>
<tr>
<td>PD</td>
<td>0.014435</td>
<td></td>
<td>0.015943</td>
<td></td>
<td>0.016947</td>
<td>0.018002</td>
</tr>
<tr>
<td>NG</td>
<td>0.011065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNR</td>
<td>0.010995</td>
<td>0.013707</td>
<td>0.013429</td>
<td>0.013383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>0.010051</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.005366</td>
</tr>
</tbody>
</table>

MO (modification) is significantly underused independent of L1
In my opinion this statement holds. The often very theoretical approach especially in Denmark where ...

...and exactly for this reason ...

Perhaps not when

Only then do they develop...

frequencies normalized per 1000 edges
modified element – results

• all categories are frequently modified in both L1 and L2
• but *all* syntactic relations possible for modification are underused
• modifiers of adverbs show the strongest underuse
If she makes her career, …
Some have success [with this] …
One can, as mentioned above …
To make money on a criminal basis …
… criminality increases steadily …
which still exists …

<table>
<thead>
<tr>
<th>func</th>
<th>L2 (norm)</th>
<th>L1 (norm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>14,6162802</td>
<td>12,8218827</td>
</tr>
<tr>
<td>PROADV</td>
<td>7,41011413</td>
<td>6,73148841</td>
</tr>
<tr>
<td>COMPARE</td>
<td>0,26343296</td>
<td>0,27475463</td>
</tr>
<tr>
<td>PREP</td>
<td>44,8600831</td>
<td>48,5857769</td>
</tr>
<tr>
<td>ADJ</td>
<td>12,7722495</td>
<td>17,5842962</td>
</tr>
<tr>
<td>ADV</td>
<td>61,8302642</td>
<td>87,7230473</td>
</tr>
</tbody>
</table>

frequencies normalized per 1000 edges
modifier – results

• categories of different complexity (lexemes to sentences) are used for modification; modification is frequent in L2 and L1
• some categories are underused by the learners, two categories are slightly overused
• adverbs and (adverbially used) adjectives show the strongest underuse

33
modification

- are the effects form-based or function-based?
  - are all adverbs underused?
    - no; auch, noch etc. overused
  - are certain adverbs (forms) underused?
    - yes
  - are certain adverbial functions underused?
    - yes
  - are certain adverbs (forms) underused in certain functions?
    - yes
  - is modification generally underused?
    - (or do learners make up for the underuse of adverbs by other means of modification?)
    - yes
summary: modification in Falko

• modification is a difficult category for learners of GFL
  – previous evidence: form-based
  – previous hypotheses: ’transfer‘, polyfunctionality

• additional syntactic evidence shows the syntactic function ’modification‘ is underused, independent of form & independent of L1 of the learners
methodological conclusions

• in annotation separation of form and function necessary

• parsing of learner data necessary to find syntactic functions

• explicit target hypotheses: making interpretation visible and learner language parsable

• multi-layer architectures
Thank you!
Merci!
Danke!

Falko:
http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko

contact: anke.luedeling@rz.hu-berlin.de
analysis of syntactic annotation: modifiers

• certain syntactic classes of adverbs are underused
• adverbs are syntactically analyzed as modifiers
• research question:
  – is adverb underuse due to lexical properties of certain adverbs?
    do learners compensate for this underuse with other means of modifications (e.g. PPs)?
  – or do learners simply underuse modifiers (of any kind) (adverb underuse would then be a result of the general underuse of modifiers)?
summary

• categorization
• additionally, there is a purely syntactic effect: MO is *structurally* underused by the learners
• why is MO difficult?
  ➢ semantics: now we would have to look at different semantic classes of modification (temporal, local, ….) – further research …
  ➢ word order (topology): placement problems in the German middle field – further research
  ➢ categorial effect: does the complexity of categories play a role?
  ➢ …
MO – summary

• the lexical ADV underuse is still visible
• additionally, there is a purely syntactic effect: MO is \textit{structurally} underused by the learners
• why is MO difficult?
  \begin{itemize}
  \item semantics: now we would have to look at different semantic classes of modification (temporal, local, …) – further research …
  \item word order (topology): placement problems in the German middle field – further research
  \item categorial effect: does the syntactic complexity of categories play a role?
  \item ...
  \end{itemize}
• in order to abstract away from semantic and word order effects we look at the vorfeld
summary

- research question: how does syntactic annotation of L2 learner data and interpretations of it help in understanding interlanguage/acquisition processes?
- interlanguage ← learner corpus
- underuse as a diagnostic for structural difficulties
- Falko
  - design: advanced learners of German, written, essays, metadata, control group
  - annotation: target hypotheses, automatic edit errors, pos, lemma, more error annotation, syntactic annotation (Berkeley parser) of target hypotheses
  - architecture: multi-layer, standoff, searchable with Annis2
summary – adverbs and modification

• from lexical studies we know that learners underuse adverbs
• modification is also generally underused

➢ combination of factors
➢ syntactic annotation helps us in finding acquisition patterns that combine lexical, categorial, topological and functional properties
interlanguage & data

• further assumption: interlanguage can be researched through the analysis of (naturally occurring) learner data
• one type of data: learner corpora
• analysis
  – error analysis
    → analysis of learner data wrt a 'correct' form
  – contrastive interlanguage analysis (CIA)
    → analysis of the learner data wrt to another corpus
background: interlanguage

• assumption: learners of a second/foreign language have a systematic internal grammar (interlanguage), different from the internal grammar of L1 speakers of the target language

• interlanguage is influenced by
  – the learners' L1 (transfer, interference)
  – the structure of the L2
  – general learning principles
  – mode of acquisition / teaching method / learning strategies

• Selinker (1972), Nickel (1998) and many others
interlanguage & data

• further assumption: interlanguage can be researched through the analysis of (naturally occurring) learner data
• one type of data: learner corpora
• analysis
  – error analysis
    → analysis of learner data wrt a 'correct' form
  – contrastive interlanguage analysis (CIA)
    → analysis of the learner data wrt to another corpus
comparison of sentence length

![Sentence length comparison](image)
data used in the study XXX

- Falko subcorpus
- the largest L1 groups (da: Danish, en: English, fr: French, pl: Polish, rz: Russian)
- 58210 tokens of too small L1s groups (pre-hoc control)

<table>
<thead>
<tr>
<th>German</th>
<th>L2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>de 88736</td>
<td>da 15593</td>
<td></td>
</tr>
<tr>
<td>en 21600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fr 7786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pl 18100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ru 11203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88736</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total 163018</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
grammatical function in the vorfeld: subject

<table>
<thead>
<tr>
<th>func_cat/pos</th>
<th>L1 (norm)</th>
<th>L2 (norm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SB</td>
<td>185,4354354</td>
<td>222,0760826</td>
</tr>
<tr>
<td>SB_PPER</td>
<td>84,45945946</td>
<td>120,4977742</td>
</tr>
<tr>
<td>SB_PDS</td>
<td>27,4024024</td>
<td>29,54269527</td>
</tr>
<tr>
<td>SB_PIS</td>
<td>29,27927928</td>
<td>28,73330635</td>
</tr>
<tr>
<td>SB_NN</td>
<td>11,07357357</td>
<td>11,33144476</td>
</tr>
<tr>
<td>SB_NE</td>
<td>6,569069069</td>
<td>4,552812626</td>
</tr>
<tr>
<td>SB_PWS</td>
<td>3,753753754</td>
<td>3,439902873</td>
</tr>
<tr>
<td>SB_NP</td>
<td>142,454955</td>
<td>169,3646297</td>
</tr>
</tbody>
</table>

comparison of different subjects in L2 and L1 frequencies normalized per 1000 main clauses
aside: Annis

- search window
- match count
- corpus selection
- metadata corpus
- metadata text
aside: Annis
aside: Annis

nur noch
only still
just

MOVS = MOVEDsource
tokens are reordered

MOVT = MOVEDtarget
token should appear here

tokens in complete text

http://korpling.german.hu-berlin.de/falko-suche
## Parser Evaluation on L1/L2

Evaluation of constituent structure with GF labels (evalb)

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
<th>Tagging acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>63.93</td>
<td>64.27</td>
<td>64.10</td>
<td>91.93</td>
</tr>
<tr>
<td>L2</td>
<td>68.16</td>
<td>69.44</td>
<td>68.79</td>
<td>92.85</td>
</tr>
<tr>
<td>Tiger*</td>
<td>69.23</td>
<td>70.41</td>
<td>69.81</td>
<td>52</td>
</tr>
</tbody>
</table>

---

Tiger* Berkeley results on the Tiger Treebank (Petrov & Klein, 2008)
vorfeld

• it is often assumed that in German only one constituent is allowed before the finite verb (V2-constraint, vorfeld-constraint)
• the vorfeld is often studied in learner language (indication of advancedness, information structure)
• in Falko: there is no significant difference in the vorfeld complexity between L1 and L2 – but do learners and native speakers use the same elements in the vorfeld?

➢ combination of topological information, functional information and categorial information

• Haberzettl (1998), Walter, Doolittle & Schmidt (2007)
elements in the vorfeld (independent of function)

<table>
<thead>
<tr>
<th>cat/pos</th>
<th>L1 (norm)</th>
<th>L2 (norm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>152,965465</td>
<td>179,481991</td>
</tr>
<tr>
<td>NN</td>
<td>10,3228228</td>
<td>11,3314448</td>
</tr>
<tr>
<td>PP</td>
<td>63,2507508</td>
<td>90,0445164</td>
</tr>
<tr>
<td>PPER</td>
<td>96,4714715</td>
<td>130,210441</td>
</tr>
<tr>
<td>ADV</td>
<td>89,9024024</td>
<td>75,9813841</td>
</tr>
<tr>
<td>AVP</td>
<td>15,2027027</td>
<td>10,0161878</td>
</tr>
<tr>
<td>PDS</td>
<td>28,3408408</td>
<td>27,9239174</td>
</tr>
<tr>
<td>PIS</td>
<td>28,5285285</td>
<td>27,7215702</td>
</tr>
<tr>
<td>AP</td>
<td>3,19069069</td>
<td>3,43990287</td>
</tr>
</tbody>
</table>

comparison of vorfeld-elements in L2 and L1 frequencies normalized per 1000 main clauses
modifiers in the vorfeld

<table>
<thead>
<tr>
<th>Modifier</th>
<th>L2 Frequency</th>
<th>L1 Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>MO</td>
<td>207,7702703</td>
<td>223,6948604</td>
</tr>
<tr>
<td>MO_PP</td>
<td>63,43843844</td>
<td>90,95507892</td>
</tr>
<tr>
<td>MO_AVP</td>
<td>15,2027027</td>
<td>10,01618778</td>
</tr>
<tr>
<td>MO_ADV</td>
<td>91,02852853</td>
<td>76,79077297</td>
</tr>
<tr>
<td>MO_PROAV</td>
<td>21,02102102</td>
<td>29,0368272</td>
</tr>
<tr>
<td>MO_AP</td>
<td>2,064564565</td>
<td>1,214083367</td>
</tr>
<tr>
<td>MO_ADJD</td>
<td>6,193693694</td>
<td>6,475111291</td>
</tr>
<tr>
<td>MO_PWAV</td>
<td>6,193693694</td>
<td>5,261027924</td>
</tr>
</tbody>
</table>

comparison of different modifier categories in L2 and L1 frequencies normalized per 1000 main clauses
summary: modification in the vorfeld

• learners generally use modification in the vorfeld as often as the native speakers
• learners have a different distribution of elements in the vorfeld
  – they overuse PPs (although they slightly underuse modifying PPs generally)
  – they also significantly overuse simple personal pronouns
  – the same categories (adverbs, adverbial phrases) that are underused everywhere as modifiers are also underused in the vorfeld
  – learners 'compensate' this by overusing prepositional phrases and pronominal adverbs
• syntactic complexity does not seem to be the relevant category

➤ back to lexical and semantic factors ...
➤ further studies: other topological areas in the sentence
aside: annis search & statistics

- represents modified element
- represents modifier
- gets frequencies for #1 and #2
aside: Annis

- we search Falko in our freely available search tool Annis2
- multi-layer standoff model (token annotation, span annotation, graphs, pointing relations)
- search across all annotation layers

- Chiarcos et al. (2008), Zeldes et al. (2009), Zipser & Romary (2010),
  http://www.sfb632.uni-potsdam.de/d1/annis/
syntactic annotation of learner corpora for acquisition research

• many studies of syntactic phenomena in learner corpora, usually on the basis of surface structures (manually, pos tags, lexical cues etc.) for German see e.g. Diehl et al. (2000), Ahrenholtz (2008), Doolittle (2008), Breckle & Zinsmeister (submitted)

• several (very few) parsed learner corpora, often not publically available
Dickinson & Ragheb (2009), Rosén & de Smedt (to appear)
syntactic annotation of learner corpora for CALL

- parsing learner data would help in generating intelligent answers to learner errors in call systems – a lot of research in this area – usually not helpful for our research question
  - often very restricted domains (question answering, fill in the blanks exercises etc.)
  - sometimes errors are explicitly introduced into 'native' data
- another goal: making parsers robust against data errors – again not directly helpful for our research question
- still: interesting results wrt to parsing techniques / evaluation techniques etc.

learner corpora for GFL

- many learner corpora for English, more and more learner corpora for other languages

- For German very few freely available learner corpora
  - LeaP (spoken)
  - AleSKO (in construction)
  - Ursula Weinberger ("")
  - Falko
annotation of learner data: format

• many learner corpora are not annotated
• some are annotated with error tags, usually tabular formats or tree formats (XML), typically not standoff, typically not amendable by the user
• some (few) are annotated on other levels (pos, lemma etc.)
consider: *An der anderen Seite, wenn da kein Feminismus wäre, stünden wir noch nur in der Küche und köchten wir.*

(fkb034_2008_07)

~ "On the other hand, if there were no feminism, we would still only stand in the kitchen and cook."
annotation of learner data: target hypothesis

• consider: *An der anderen Seite, wenn da kein Feminismus wäre, stünden wir noch nur in der Küche und köchten wir.*

(fkb034_2008_07)

~ "On the other hand, if there were no feminism, we would still only stand in the kitchen and cook."
annotation of learner data: target hypothesis

• all error tags depend on an (at least implicit) correct version of a learner utterance → target hypothesis

• Falko: explicit target hypotheses

• often there are several ways of correcting an utterance

th1: *Auf der anderen Seite, wenn da kein Feminismus wäre, stünden wir nur noch in der Küche und kochten.*

th2: *Andererseits stünden wir, wenn es keinen Feminismus gäbe, nur noch in der Küche und kochten.*
<table>
<thead>
<tr>
<th>learner utterance</th>
<th>target hypothesis 1</th>
<th>errors</th>
<th>target hypothesis 2</th>
<th>errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>An</td>
<td>Auf</td>
<td>CHA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>der</td>
<td>der</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>anderen</td>
<td>anderen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seite</td>
<td>Seite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andererseits</td>
<td></td>
<td></td>
<td></td>
<td>MERGE</td>
</tr>
<tr>
<td>stünden</td>
<td></td>
<td></td>
<td></td>
<td>MOVT</td>
</tr>
<tr>
<td>wir</td>
<td></td>
<td></td>
<td></td>
<td>MOVT</td>
</tr>
<tr>
<td>,</td>
<td></td>
<td></td>
<td></td>
<td>INS</td>
</tr>
<tr>
<td>wenn</td>
<td>wenn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wenn</td>
<td>wenn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td>da</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>da</td>
<td>da</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kein</td>
<td>kein</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kein</td>
<td>kein</td>
<td></td>
<td></td>
<td>CHA</td>
</tr>
<tr>
<td>Feminismus</td>
<td>Feminismus</td>
<td></td>
<td>Feminismus</td>
<td></td>
</tr>
<tr>
<td>wäre</td>
<td>wäre</td>
<td></td>
<td>gäbe</td>
<td>CHA</td>
</tr>
<tr>
<td>,</td>
<td>,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>,</td>
<td>,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stünden</td>
<td>stünden</td>
<td></td>
<td></td>
<td>MOVS</td>
</tr>
<tr>
<td>wir</td>
<td>wir</td>
<td></td>
<td></td>
<td>MOVS</td>
</tr>
<tr>
<td>nur</td>
<td></td>
<td>MOVT</td>
<td>nur</td>
<td>MOVT</td>
</tr>
<tr>
<td>noch</td>
<td>noch</td>
<td></td>
<td>noch</td>
<td></td>
</tr>
<tr>
<td>nur</td>
<td></td>
<td>MOVS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>in</td>
<td>in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>der</td>
<td>der</td>
<td></td>
<td></td>
<td>der</td>
</tr>
<tr>
<td>Küche</td>
<td>Küche</td>
<td></td>
<td>Küche</td>
<td></td>
</tr>
<tr>
<td>und</td>
<td>und</td>
<td></td>
<td>und</td>
<td></td>
</tr>
</tbody>
</table>

67
how can we detect acquisition problems?

- structures that are unique for the L2 or different from the learners' L1s (transfer)
- structures that are judged to be difficult by the learners
- structures that contain many errors
- underused structures
how can we detect acquisition problems?

• structures that are unique for the L2 or different from the learners' L1s (transfer)
  ➢ grammatical analysis
  ➢ proved to be extremely problematic; no straightforward transfer
• structures that are judged to be difficult by the learners
  ➢ intuition of the learners (unsystematic, dependent on teaching)
  ➢ experiments
• structures that contain many errors
  ➢ intuition of the teachers (unsystematic)
  ➢ corpus analysis, error analysis
• underused structures
  ➢ corpus analysis, Contrastive Interlanguage Analysis
how can we detect acquisition problems?

• structures that are unique for the L2 or different from the learners' L1s (transfer)
  ➢ grammatical analysis
  ➢ proved to be extremely problematic; no straightforward transfer
• structures that are judged to be difficult by the learners
  ➢ intuition of the learners (unsystematic, dependent on teaching)
  ➢ experiments
• structures that contain many errors
  ➢ intuition of the teachers (unsystematic)
• underused structures
parser evaluation

- evaluation of constituent structure (evalb)

<table>
<thead>
<tr>
<th>&gt;40</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
<th>Tagging acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>73.61</td>
<td>74.00</td>
<td>73.80</td>
<td>91.93</td>
</tr>
<tr>
<td>L2</td>
<td>77.59</td>
<td>79.04</td>
<td>78.31</td>
<td>92.85</td>
</tr>
<tr>
<td>Negra*</td>
<td>80.01</td>
<td>80.01</td>
<td>80.01</td>
<td></td>
</tr>
</tbody>
</table>

- L2 easier to parse than L1
- possible reasons: sentence length / L1 syntactic structure might be more complex
  - we can use parser output to compare L1 and L2
  - *Berkeley results on the Negra Treebank (Petrov & Klein, 2007)
exemplary study about modifiers

• further probes into the adverb underuse
  – underuse statistics of syntactic categories: types of modification
  – underuse statistics & a combined search over positions (fields), categories and functions (vorfeld)
underuse of adverb chains

• the syntactic adverb classes were (manually) annotated (in essence this is a more fine-grained pos categorization)
• many studies about adverbs in learner language – analysis purely lexical
• the different distributions suggest that syntax might be relevant for understanding learner language
• however, the syntactic information codable at token level is too limited: we need hierarchical relations (dependencies, constituents)
• Möllering (2004), Vyatkina (2007) etc.