Thoughts on Learner Data and Dependency Parsing

Niels Ott, Ramon Ziai, Julia Krivanek, Detmar Meurers
Universität Tübingen
SFB 833, Project A4

Second Tübingen-Berlin Meeting on Analyzing Learner Language
Tübingen, 5./6. Dezember 2011
Overview

Introduction and Motivation

Learner Language and Dependency Annotation

Approximated Target Hypotheses

Rule-Based vs. Data-Driven
  Hypothesis
  Corpora used
  Parsers used
  Overall Results
  Drop between UAS & LAS
  Results by dependency type
  A subjectless example

Conclusion
General Motivation

Why dependency parsing?

- Focus on lexical dependency structure as an interface to interpretation. → CoMiC project compares meaning of answers to reading comprehension questions
- At the same time, to characterize the nature of learner language, capturing morphosyntactic dependencies can also be an important goal (→ SLA research)
The CoMiC project investigates how the meaning of student answers can be compared to the meaning of target answers in reading comprehension exercises.

Data: Corpus from German classes in the US, Ohio State University (Prof. Kathryn Corl), University of Kansas (prof. Nina Vyatkina).

Target answers and student answers are compared with respect to meaning, not form.

- Trying to detect automatically: Did the student answer the question correctly or not?

We want to parse German learner language automatically with dependency parsers.

These data are not annotated with errors or target hypotheses.
Our experimental system CoMiC-DE performs meaning comparison on many levels, beginning from simple token overlap.

So far, our most sophisticated level of linguistic representation is based on Lexical Resource Semantics (LRS, Richter & Sailer 2003).

Hahn & Meurers (2011) present an approach to the construction LRS representation from dependency structures.

Naturally, we need well-behaved dependency structures of our learner data in order to construct good LRS representations.

Furthermore, we use dependency triples directly in the system.
Ott & Ziai (2010) trained a statistical dependency parser on a dependency-converted version of the Tüba-D/Z treebank and used it to parse learner language.

- CREG109, data set with 109 manually student answers
- We are currently working on an extended data set containing more data and questions and target answers.

Annotation scheme used: the one described by Foth (2006).
The dependency annotation scheme by Foth (2006) has not been designed for learner language.

Hence, we are using an annotation scheme that simply cannot handle many constructions in the learner data.

What are possible solutions to this issue?

1. Annotating interlanguage as a system in its own right using a special annotation scheme (Dickinson & Ragheb 2009).

2. Annotating target hypotheses that map to well-formed language and annotate these (or parse: Hirschmann et al. 2010).
Annotating/Parsing Interlanguage

- So, if we would stick to interlanguage as a system in its own right?
- Interlanguage is influenced by many learner-dependent factors (stage of acquisition, L1, background, etc).
  - Difficult to capture in a general parsing model.
Aside: Isn’t it only a robustness issue?

- Arguably, robust tools should be able to deal with learner language to some extent.
- Foster (2007) automatically ‘damaged’ the Penn Treebank with simulated learner errors and trained a parser on it to achieve more error-tolerance.
- Still, this does not solve the problem of abusing an annotation scheme.
- Possibly, there is a difference between learner levels
  - Very advanced learners will be close to native speakers, so using native language categories might still be OK.
  - In our data, we have many beginners and intermediate learners that produce language that often is impossible to treat with native language categories.
- Robustness is good for us but robustness alone does not help us.
Annotating/Parsing Target Hypotheses

- Using target hypotheses seems appealing in our situation.
  - Standard NLP tools could be used in the tool chain, since we would have well-behaved language back again.
  - Still, we would have an explicit mapping back to the original learner data.
- However, we do not want to annotate target hypotheses.
  - Our corpus is large, it would be a lot of work.
  - It would not be applicable to tutoring systems that aim at giving feedback on unseen learner data.
  - Target hypotheses in the sense of Falko’s ZH1 (Reznicek et al. 2010) would be great, but also they would provide more than we need.
Let’s do it Automatically?!

Can we create target hypotheses automatically?

► It seems quite impossible, since humans put a lot of thinking into creating these.
► But perhaps we can create **approximated target hypotheses**, that provide enough wellformedness for our tools.
Example: Missing Verbs (I)
(From Ott & Ziai 2010)

33,9 Prozent, die über 25 Jahre alt [sind], sind Männer.
33.9 percent, who over 25 years old [are], are men.

→ Without the verb as a functor, it is hard for annotators and parsers to attach the dependency relations to a token for which it makes sense.
Example: Missing Verbs (II)

- Missing verbs let us lose the main functor-argument relations in the sentence.
- Detecting missing verbs and inserting a **dummy verb** could help.
  - This dummy verb would have no lexical form and no meaning.
  - Parsers would have to be trained to work properly with such underspecified tokens.
- A dummy verb would still allow for an almost-complete analysis, also in the automatically constructed LRS representation!
Parsing with Approximated Target Hypotheses

- A component for constructing approximated target hypotheses **needs to be invented**.
  - As a starting point, we can focus on missing tokens, this will be difficult enough.
  - Insertion of **dummy tokens that provide function, but not meaning**.

- Inspired by Foster (2007), we could systematically sprinkle these dummy tokens into a treebank such as Tüba-D/Z in order to train a **robust parser** to work on it.

- Parsing approximated target hypotheses with a native language dependency scheme, combined with robustness.
Comparing Rule-based and Data-Driven Parsing of Learner Language (Krivanek & Meurers 2011)

- Two dependency parsing approaches
  - **Rule-based**: WCDG Parser (Foth & Menzel 2006)
  - **Data-driven**: MaltParser (Nivre et al. 2007)
Comparing Rule-based and Data-Driven Parsing of Learner Language (Krivanek & Meurers 2011)

- Two dependency parsing approaches
  - Rule-based: WCDG Parser (Foth & Menzel 2006)
  - Data-driven: MaltParser (Nivre et al. 2007)

- Two test corpora:
  - Native language: German dependency treebank derived from TüBa-D/Z (Telljohann et al. 2004)
  - Learner language: CREG-109 (Ott & Ziai 2010)
Comparing Rule-based and Data-Driven Parsing of Learner Language
(Krivanek & Meurers 2011)

- Two dependency parsing approaches
  - Rule-based: WCDG Parser (Foth & Menzel 2006)
  - Data-driven: MaltParser (Nivre et al. 2007)

- Two test corpora:
  - Native language: German dependency treebank derived from TüBa-D/Z (Telljohann et al. 2004)
  - Learner language: CREG-109 (Ott & Ziai 2010)

- Two types of grammatical relations:
  - Argument relations: obligatory, linguistic knowledge
  - Adjunct relations: optional, relevance of world knowledge
Hypothesis about impact of parsing method

- Rule-based WCDG relies on a hand-written lexicon.
  - Hand-written, information-rich lexicon encodes knowledge about linguistic relations.
Hypothesis about impact of parsing method

- Rule-based WCDG relies on a hand-written lexicon.
  - Hand-written, information-rich lexicon encodes knowledge about linguistic relations.
  
  → **Rule-based** approach will fare better in identifying the **core functor-argument relations**.
Hypothesis about impact of parsing method

- Rule-based WCDG relies on a hand-written lexicon.
  - Hand-written, information-rich lexicon encodes knowledge about linguistic relations.
    -> **Rule-based** approach will fare better in identifying the core functor-argument relations.

- The statistical MaltParser is trained on annotated corpora.
  - Corpora encode a combination of language competence, language use, and facts about the world.
Hypothesis about impact of parsing method

- Rule-based WCDG relies on a hand-written lexicon.
  - Hand-written, information-rich lexicon encodes knowledge about linguistic relations.
  - **Rule-based** approach will fare better in identifying the core functor-argument relations.

- The statistical MaltParser is trained on annotated corpora.
  - Corpora encode a combination of language competence, language use, and facts about the world.
  - **Data-driven** approach will fare better in identifying adjunct relations.

- Test hypotheses in parsing experiments.
Thoughts on Learner Data and Dependency Parsing
Niels Ott, Ramon Ziai, Julia Krivanek, Detmar Meurers

Introduction and Motivation
Learner Language and Dependency Annotation
Approximated Target Hypotheses
Rule-Based vs. Data-Driven Hypothesis

Corpora used: Learner corpus CREG (Meurers, Ott & Ziai 2010)

- Corpus of Reading Comprehension Exercises in German
  - answers to reading comprehension questions
  - written by US college students at the beginner and intermediate levels of German programs

Parsing experiments

Corpora used: Learner corpus CREG (Meurers, Ott & Ziai 2010)

- Corpus of Reading Comprehension Exercises in German
  - answers to reading comprehension questions
  - written by US college students at the beginner and intermediate levels of German programs
Parsing experiments

Corpora used: Learner corpus CREG (Meurers, Ott & Ziai 2010)

- Corpus of Reading Comprehension Exercises in German
  - answers to reading comprehension questions
  - written by US college students at the beginner and intermediate levels of German programs

- CREG-109 is a subset manually annotated using dependency annotation scheme of Foth (2006).
  - 109 sentences (sentence length: avg. 8.26, max. 17)
  - 17 of those are ungrammatical
    - errors in word order, agreement, and case government
    - dependencies were annotated on a grammatical target hypothesis (with lexical mappings to the learner tokens)
CREG Example

T:  
- (Reading comprehension text)

Q:  Was sind die Kritikpunkte, die Leute über Hamburg äußern?

‘What are the objections people have about Hamburg?’

TA: Der Gestank von Fisch und Schiffsdiesel an den Kais.
The stink of fish and fuel on the quays.

SA: Der Geruch von Fish und Schiffsdiesel beim Hafen.
The smell of fish and fuel at the port.
Parsing experiments
Corpora used: Native language corpus

- TüBa-D/Z treebank (Ver. 5; Telljohann et al. 2004)
  - newspaper text
  - 10 % (4142 sentences) as test set
  - 90 % used as training set for MaltParser
Parsing experiments
Corpora used: Native language corpus

- TüBa-D/Z treebank (Ver. 5; Telljohann et al. 2004)
  - newspaper text
  - 10 % (4142 sentences) as test set
  - 90 % used as training set for MaltParser

- For comparability, both CREG-109 and TüBa-D/Z test corpora were automatically pos-tagged with TnT tagger (Brants 2000) using the STTS tagset (Thielen et al. 1999)
Parsers used

- MaltParser: state-of-the-art system for transition-based dependency parsing (Nivre et al. 2007)
Parsers used

- MaltParser: state-of-the-art system for transition-based dependency parsing (Nivre et al. 2007)
- WCDG: Weighted Constraint Dependency parsing for German (Foth et al. 2005; Foth & Menzel 2006)
  - hand-written grammar, with some other components:
    - heuristic search option (*frobbing*)
    - some stochastic predictor components (chunker, supertagger, probabilistic shift-reduce parser)
    - efficiency remains an issue
### Overall Scores (using eval.pl, Buchholz & Marsi 2006)

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WCDG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
</tr>
<tr>
<td><strong>MaltParser</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>84.04%</td>
<td>87.25%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>78.12%</td>
<td>84.56%</td>
</tr>
</tbody>
</table>

- overall results quite similar for both parsers
## Overall Scores (using eval.pl, Buchholz & Marsi 2006)

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WCDG</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
</tr>
<tr>
<td><strong>MaltParser</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>84.04%</td>
<td>87.25%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>78.12%</td>
<td>84.56%</td>
</tr>
</tbody>
</table>

- overall results quite similar for both parsers
  - WCDG: robust parsing of learner language
Overall Scores (using eval.pl, Buchholz & Marsi 2006)

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
</tr>
<tr>
<td>WCDG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MaltParser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>84.04%</td>
<td>87.25%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>78.12%</td>
<td>84.56%</td>
</tr>
</tbody>
</table>

- overall results quite similar for both parsers
  - WCDG: robust parsing of learner language
  - MALT better for native language (which it was trained on)
## Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>WCDG</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LAS</td>
<td>UAS</td>
<td>UAS – LAS</td>
<td></td>
</tr>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
<td></td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
<td>= 2.79% diff.</td>
</tr>
</tbody>
</table>

- 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native).
- Parallel observation also holds for Maltparser results.
- Hypothesis: This 2.79% higher drop for CREG-109 may result from ungrammatical learner sentences.
- Manually inspected the 53 cases (7.08%) where the parser assigned correct relations but false labels.
- 21 resulted from ungrammaticality = 2.8%!
Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th>WCDG</th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
</tr>
</tbody>
</table>

= 2.79% diff.

▶ 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
## Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td><strong>4.29%</strong></td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td><strong>7.08%</strong> = <strong>2.79%</strong> diff.</td>
</tr>
</tbody>
</table>

- 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
Thoughts on Learner Data and Dependency Parsing

Niels Ott, Ramon Ziai, Julia Krivanek, Detmar Meurers

Introduction and Motivation

Learner Language and Dependency Annotation

Approximated Target Hypotheses

Rule-Based vs. Data-Driven Hypothesis

Corpora used

Parsers used

Overall Results

Drop between UAS & LAS

Results by dependency type

A subjectless example

Conclusion

References

Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th>WCDG</th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
</tr>
</tbody>
</table>

▶ 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)

▶ parallel observation also holds for Maltparser results

▶ 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)

▶ parallel observation also holds for Maltparser results
Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
</tr>
</tbody>
</table>

- 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
  - parallel observation also holds for Maltparser results
- Hypothesis: This 2.79% higher drop for CREG-109 may result from ungrammatical learner sentences.
Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
</tr>
</tbody>
</table>

- 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
  - parallel observation also holds for Maltparser results
- Hypothesis: This 2.79% higher drop for CREG-109 may result from ungrammatical learner sentences.
  - Manually inspected the 53 cases (7.08%) where the parser assigned correct relations but false labels
Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08% = 2.79% diff.</td>
</tr>
</tbody>
</table>

- 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
  - parallel observation also holds for Maltparser results
- Hypothesis: This 2.79% higher drop for CREG-109 may result from ungrammatical learner sentences.
  - Manually inspected the 53 cases (7.08%) where the parser assigned correct relations but false labels
  - 21 resulted from ungrammaticality
Investigating the drop from UAS to LAS

<table>
<thead>
<tr>
<th></th>
<th>LAS</th>
<th>UAS</th>
<th>UAS – LAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TüBa-D/Z</td>
<td>81.42%</td>
<td>85.71%</td>
<td>4.29%</td>
</tr>
<tr>
<td>CREG-109</td>
<td>79.28%</td>
<td>86.36%</td>
<td>7.08%</td>
</tr>
</tbody>
</table>

▶ 7.08% drop from UAS to LAS for CREG-109 (learner), but only 4.29% drop for TüBa-D/Z (native)
  ▶ parallel observation also holds for Maltparser results

▶ Hypothesis: This 2.79% higher drop for CREG-109 may result from ungrammatical learner sentences.
  ▶ Manually inspected the 53 cases (7.08%) where the parser assigned correct relations but false labels
  ▶ 21 resulted from ungrammaticality = 2.8%! 

≈ 2.79% diff.
Investigating the drop between UAS and LAS
Example: mixed up subject & object due to agreement error

(1) *Seine Eltern hat BA geholfen.*

his_{pl} parents_{pl} has_{sg} BA_{sg} helped

intended: His parents have helped BA.

[WCDG parse]
## CREG-109 results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>MaltParser</th>
<th>WCDG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Prec.</td>
</tr>
<tr>
<td><strong>Argument relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>95</td>
<td>84.21</td>
</tr>
<tr>
<td>OBJA</td>
<td>52</td>
<td>65.38</td>
</tr>
<tr>
<td>PRED</td>
<td>26</td>
<td>61.54</td>
</tr>
<tr>
<td>AUX</td>
<td>23</td>
<td>60.87</td>
</tr>
<tr>
<td><strong>Modifier relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>44</td>
<td>65.91</td>
</tr>
<tr>
<td>PP</td>
<td>32</td>
<td>75.00</td>
</tr>
</tbody>
</table>
### CREG-109 results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>#</th>
<th>MaltParser</th>
<th>WCDG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recall</td>
<td>Prec.</td>
</tr>
<tr>
<td>Argument relations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>95</td>
<td>84.21</td>
<td>80.00</td>
</tr>
<tr>
<td>OBJA</td>
<td>52</td>
<td>65.38</td>
<td>70.83</td>
</tr>
<tr>
<td>PRED</td>
<td>26</td>
<td>61.54</td>
<td>69.57</td>
</tr>
<tr>
<td>AUX</td>
<td>23</td>
<td>60.87</td>
<td>87.50</td>
</tr>
<tr>
<td>Modifier relations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>44</td>
<td>65.91</td>
<td>56.86</td>
</tr>
<tr>
<td>PP</td>
<td>32</td>
<td>75.00</td>
<td>55.81</td>
</tr>
</tbody>
</table>
## CREG-109 results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>MaltParser</th>
<th>WCDG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recall</td>
<td>Prec.</td>
</tr>
<tr>
<td><strong>Argument relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>84.21</td>
<td>80.00</td>
</tr>
<tr>
<td>OBJA</td>
<td>65.38</td>
<td>70.83</td>
</tr>
<tr>
<td>PRED</td>
<td>61.54</td>
<td>69.57</td>
</tr>
<tr>
<td>AUX</td>
<td>60.87</td>
<td>87.50</td>
</tr>
<tr>
<td><strong>Modifier relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>65.91</td>
<td>56.86</td>
</tr>
<tr>
<td>PP</td>
<td>75.00</td>
<td>55.81</td>
</tr>
</tbody>
</table>
TüBa-D/Z results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>MaltParser</th>
<th></th>
<th>WCDG</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>Recall</td>
<td>Prec.</td>
<td>Recall</td>
</tr>
<tr>
<td><strong>Argument relations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>5408</td>
<td>83.54</td>
<td>87.05</td>
<td>89.00</td>
</tr>
<tr>
<td>OBJA</td>
<td>2658</td>
<td>75.43</td>
<td>72.96</td>
<td>79.83</td>
</tr>
<tr>
<td>PRED</td>
<td>1044</td>
<td>66.48</td>
<td>71.77</td>
<td>60.82</td>
</tr>
<tr>
<td>AUX</td>
<td>2236</td>
<td>85.73</td>
<td>89.41</td>
<td>91.77</td>
</tr>
<tr>
<td><strong>Modifier relations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>5115</td>
<td>78.92</td>
<td>77.78</td>
<td>69.72</td>
</tr>
<tr>
<td>PP</td>
<td>5562</td>
<td>71.88</td>
<td>72.26</td>
<td>69.67</td>
</tr>
</tbody>
</table>

Introduction and Motivation

Learner Language and Dependency Annotation

Approximated Target Hypotheses

Rule-Based vs. Data-Driven

Hypothesis

Corpora used

Parsers used

Overall Results

Drop between UAS & LAS

Results by dependency type

A subjectless example

Conclusion

References
### TüBa-D/Z results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>MaltParser</th>
<th>WCDG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>Recall</td>
</tr>
<tr>
<td><strong>Argument relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>5408</td>
<td>83.54</td>
</tr>
<tr>
<td>OBJA</td>
<td>2658</td>
<td>75.43</td>
</tr>
<tr>
<td>PRED</td>
<td>1044</td>
<td><strong>66.48</strong></td>
</tr>
<tr>
<td>AUX</td>
<td>2236</td>
<td>85.73</td>
</tr>
<tr>
<td><strong>Modifier relations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>5115</td>
<td><strong>78.92</strong></td>
</tr>
<tr>
<td>PP</td>
<td>5562</td>
<td><strong>71.88</strong></td>
</tr>
</tbody>
</table>
TüBa-D/Z results by dependency types

<table>
<thead>
<tr>
<th>Label</th>
<th>#</th>
<th>MaltParser</th>
<th></th>
<th></th>
<th>WCDG</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recall</td>
<td>Prec.</td>
<td>Recall</td>
<td>Prec.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argument relations</td>
<td></td>
<td>MaltParser</td>
<td>WCDG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUBJ</td>
<td>5408</td>
<td>83.54</td>
<td>87.05</td>
<td>89.00</td>
<td>89.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OBJA</td>
<td>2658</td>
<td>75.43</td>
<td>72.96</td>
<td>79.83</td>
<td>82.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRED</td>
<td>1044</td>
<td><strong>66.48</strong></td>
<td>71.77</td>
<td>60.82</td>
<td>76.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUX</td>
<td>2236</td>
<td>85.73</td>
<td>89.41</td>
<td><strong>91.77</strong></td>
<td><strong>96.11</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifier relations</td>
<td></td>
<td>MaltParser</td>
<td>WCDG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADV</td>
<td>5115</td>
<td><strong>78.92</strong></td>
<td><strong>77.78</strong></td>
<td>69.72</td>
<td>64.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td>5562</td>
<td><strong>71.88</strong></td>
<td><strong>72.26</strong></td>
<td>69.67</td>
<td>62.92</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
WCDG: robust parsing of subjectless sentences

(2) Vielleicht adoptieren ein Kind.
    perhaps adopt_{plur} a child_{sing}
(2) *Vielleicht adoptieren ein Kind.*
perhaps adopt_{plur} a child_{sing}

**MaltParser**: object attached as subject
(2) *Vielleicht adoptieren ein Kind.*
   perhaps adopt<sub>plur</sub> a child<sub>sing</sub>

**MaltParser:** object attached as subject

**WCDG parser:** subjectless analysis
Conclusion I

- Strengths of different information included in parser:
  - hand-written lexical information (WCDG) helpful in identifying **obligatory functor-argument relations**
  - world knowledge in corpora helps data-driven parsers perform well on **optional adjunct relations**

- For parsing learner language, several levels of dependency analysis are probably needed
  - robust analysis glossing over learner language specifics, as a step towards meaning
    How can one make the target hypotheses explicit that underlies this, in an automated process from data to interpretation?
  - analysis identifying specific learner language evidence
Concerning robust parsing of learner language, we will
▶ annotate a sample of CREG with target hypotheses
▶ investigate to what extent an approximated target hypothesis could be generated automatically
▶ ask the Berlin crowd how annotation on manual target hypotheses can be mapped back to learner language.
References


