Comparing Link Grammars and Dependency Grammars for parsing German histological reports

Julian Dörenberg\textsuperscript{1, 2}

julian.doerenberg@rwth-aachen.de, jdoerenberg@ukaachen.de

\textsuperscript{1} RWTH Aachen University
\textsuperscript{2} University Hospital RWTH Aachen
Secondary usage of data

- Diagnostics
- Anamnesis
- Clinical outcomes
- Treatment (doctor, nurse, ...)
- Research (e.g. statistics, AI, ...)

Histological report is floating text.

Currently done by human

Computer needs structured data.

Slide 2
Gutachten (Excerpt)

„(…)

Unterlappen mit einem 6,5 cm großen mäßiggradig differenzierten Plattenepithelkarzinom.

Minimale Randabstände des Plattenepithelkarzinoms:
- Zum Bronchusresektionsrand: 0,4 cm
- Zur Pleura visceralis: 0,1 cm
- Zum chirurgischen Resektionsrand: 0,7 cm

Nebenbefundlich abszedierende Retentionspneumonie und fibrosierte Pleura visceralis.

(...)“
Problem and example sentence

Human reads:
„Unterlappen mit einem 6,5 cm großen mäßiggradig differenzierten Plattenepithelkarzinom.“

-> Plattenepithelkarzinom is mäßiggradig differenziert
-> Plattenepithelkarzinom is 6.5cm of size
-> Plattenepithelkarzinom is located at Unterlappen (lower lobe of the lung)

Computer reads:
„Word Word Word Word Word Word Word Word Word Word.“

-> No structured information
1.) **Semantics**: Identify relevant words and define a mapping to a data format
   ⇒ Interesting research question, but not part of this talk

2.) **Syntax**: Identify grammatical relations between words
   ⇒ Use grammar and parser

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Type of Carcinoma</th>
<th>Size</th>
<th>Degree of Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>12345</td>
<td>Plattenepithelkarzinom</td>
<td>6,5cm</td>
<td>mäßiggradig</td>
</tr>
</tbody>
</table>
Link Grammars [1, 2]

- Grammar is dictionary mapping words to grammatical contexts
  
  \[\text{Karzinom} \rightarrow d_{11} \lor d_{12} \lor d_{13} \ldots\]
  \[\text{groß} \rightarrow d_{21} \lor d_{22} \lor d_{23} \ldots\]

- Parsing is satisfiability checking while complying to meta-rules rules
  - Input: Sentence and grammar
  - Output: Relation graph
  - Idea: Recursively parse regions
    - Base case: Add single relation
    - Backpropagation: Make sure to comply with four rules
  - Runtime: Exponential, but can be reduced to polynomial by using memoization

- Training algorithm is unsupervised
  - Input: Sentences including wordclasses of words
  - Output: Dictionary
  - Idea: Iteratively extend the grammar
    - Generate all possible relation graphs and evaluate for best one by using a metric called membership value
    - Relation graphs must comply to meta-rules
  - Runtime: Polynomial
(Dis)advantages of LGs

• Advantages:
  • Simple, quite intuitive formalism
  • Frequently used in information extraction for English medical reports
  • Unsupervised training algorithm only needs simple data annotations

• Disadvantages:
  ✓ Adaptations of formalism necessary for German [1]
  ✓ No parser available
  ✓ No training algorithm available
  ✓ No training data available
  • Lexicalized, unable to handle unknown words and typos
  • According to S. Kübler: Bad linguistic motivation for German
  • Can cause cycles in relation graph (not supported by linguistics)
Dependency Grammars (DGs) [3]

• Grammar is trained Recurrent Neural Net (RNN) with word embeddings

• Parsing is applying RNN operations
  • Input: Embedded words
  • Output: Relation graph tree
  • Idea: RNNs
  • Runtime: Polynomial

• Training algorithm is optimizing weights – e.g. Adam
  • Input: RNN model
  • Output: Optimized weights
  • Idea: Various
  • Runtime: Various
(Dis)advantages of DGs

**Advantages:**
- Neural parser available (Supar [4])
  - Supports different models
  - Word embeddings support unknown words and typos [6]
- Relation graph is a tree, as supported by linguistics
- Training data for German available online

**Disadvantages:**
- Rarely used for medical applications, especially German
- In literature poor performance for German [5]
- Neural Net: More training data needed, hidden states/black box, computationally expensive
- More complex annotation of training data: Link Grammar just needs word classes, Dependency Grammar needs parsing trees

Grammatical role of word:
- nominal subject (nsubj)
- numerical modifier (nummod)
- adjectival modifier (amod)
- adverbial modifier (advmod)
LGs vs DGs in practical usage

• Lexicalization := Grammar contains dictionary and cannot handle words which are not contained
  • LGs: Yes
  • DGs: No – Use word embeddings

• Public training data available
  • LGs: No
  • DGs: Yes – Universal Dependencies project [3]

• Public parser framework for German available
  • LGs: No
  • DGs: Yes – Supar [4]
Evaluation data for a DG parser

• Parser pretrained by Timothy Dozat and Christopher Manning [6]

• Training data:
  • Newspaper articles and google reviews from Universal Dependencies project [3, 6]
  • Contain twelve languages: Bulgarian, Catalan, Czech, German, English, Spanish, French, Italian, Dutch, Norwegian, Romanian/Moldavian* and Russian

• Evaluation data:
  • 200 sentences from breast biopsy reports
  • Written by two senior pathologists and annotated by me

* Using the ISO 639-1 code it remains ambiguous whether the data include sentences in Romanian, Moldavian or both and Dozat and Manning did not elaborate on that in their paper.
Evaluation concept for a DG parser

- Following idea by Gómez-Rodríguez et al. [12]

1) Unlabelled Attachment Score - UAS
   - Proportion of correctly extracted grammatical relations between words

2) Labelled Accuracy – LA
   - Proportion of correctly labelled words

3) Labelled Accuracy Score – LAS
   - Proportion of correctly labelled word with correct relation to father node

4) Application-specific:
   - Medical terms are unknown in training
     - Proportion Of correct {2,3,4}-ary relations containing at least one medical term
**Evaluation of the DG parser**

Evaluation data: 200 sentences randomly selected from the breast biopsy reports

-> Localization sentence (local.) do not have unambiguous parsing

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unlabelled Attachment Score (UAS)</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Labelled Accuracy Score (LA)</td>
<td>0.92</td>
<td>0.95</td>
</tr>
<tr>
<td>Labelled Attachment Score (LAS)</td>
<td>0.9</td>
<td>0.93</td>
</tr>
<tr>
<td>Proportion of correctly extracted 2-ary relations</td>
<td>0.95</td>
<td>0.97</td>
</tr>
<tr>
<td>Proportion of correctly extracted 3-ary relations</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>Proportion of correctly extracted 4-ary relations</td>
<td>0.88</td>
<td>0.89</td>
</tr>
</tbody>
</table>

**UAS:** Proportion of correct relations  
**LA:** Proportion of correct tags  
**LAS:** Proportion of correct tags and relations

**\{2,3,4\}-ary relations:** Proportion of correctly extracted relations containing at least one medical word
Limitations of DG parsing

• Multiword Expressions (MWEs): At least two words form a semantical unit
  • E.g. Carcinoma in situ

• No MWE parsed correctly in 200 breast biopsy reports, but corpus too small for more detailed analysis

⇒ Training data from histological reports required to analyse and enhance the performance of the DG parser on histological reports

• Nevertheless, Evaluation in real application interesting due to good performance
## Complete approach: Evaluation

<table>
<thead>
<tr>
<th>Number of HCCs</th>
<th>Fibrosis</th>
<th>Vascular invasion</th>
<th>Tumor diameter</th>
<th>Inflamm. degree</th>
<th>Distance to resection area</th>
<th>Desmet stage</th>
<th>Steatosis</th>
<th>Cirrhosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>1,4cm</td>
<td>TRUE</td>
<td>1mm</td>
<td></td>
<td>TRUE</td>
<td>TRUE</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>5,5cm</td>
<td>TRUE</td>
<td>0,3cm</td>
<td></td>
<td>TRUE</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>4,2cm</td>
<td>TRUE</td>
<td>0,3cm</td>
<td></td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>8,5cm</td>
<td>TRUE</td>
<td>3</td>
<td></td>
<td>TRUE</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>16cm</td>
<td>TRUE</td>
<td>0,1cm</td>
<td></td>
<td>TRUE</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>4,2cm</td>
<td>TRUE</td>
<td>1,5mm</td>
<td></td>
<td>TRUE</td>
<td>FALSE</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>9,5cm</td>
<td>TRUE</td>
<td>1cm</td>
<td></td>
<td>TRUE</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>8,5cm</td>
<td>TRUE</td>
<td></td>
<td></td>
<td>TRUE</td>
<td>False</td>
</tr>
<tr>
<td>1</td>
<td>TRUE</td>
<td>TRUE</td>
<td>3,6cm</td>
<td>TRUE</td>
<td>0,2cm</td>
<td>1-2</td>
<td>TRUE</td>
<td>False</td>
</tr>
</tbody>
</table>

### Problematic sentence:

(…) mit milder entzündlicher Aktivität und portaler sowie septenbildender Fibrose mit Architekturstörung (Grad 2, Stadium 3 nach Desmet).
Summary

- DGs are superior over LGs in practical usage
  - No public framework available for LGs
  - No public training data available for LGs
  - LGs are lexicalized

- Performance of LGs on histological reports remains unknown
- DG parsing performance on histological reports is very good even though no medical training data were used

⇒ Need for analysis of LG parser performance on histological reports is questionable and not recommended, use DGs instead

⇒ DG-annotated training data from histological reports required for improvement of parsing performance


