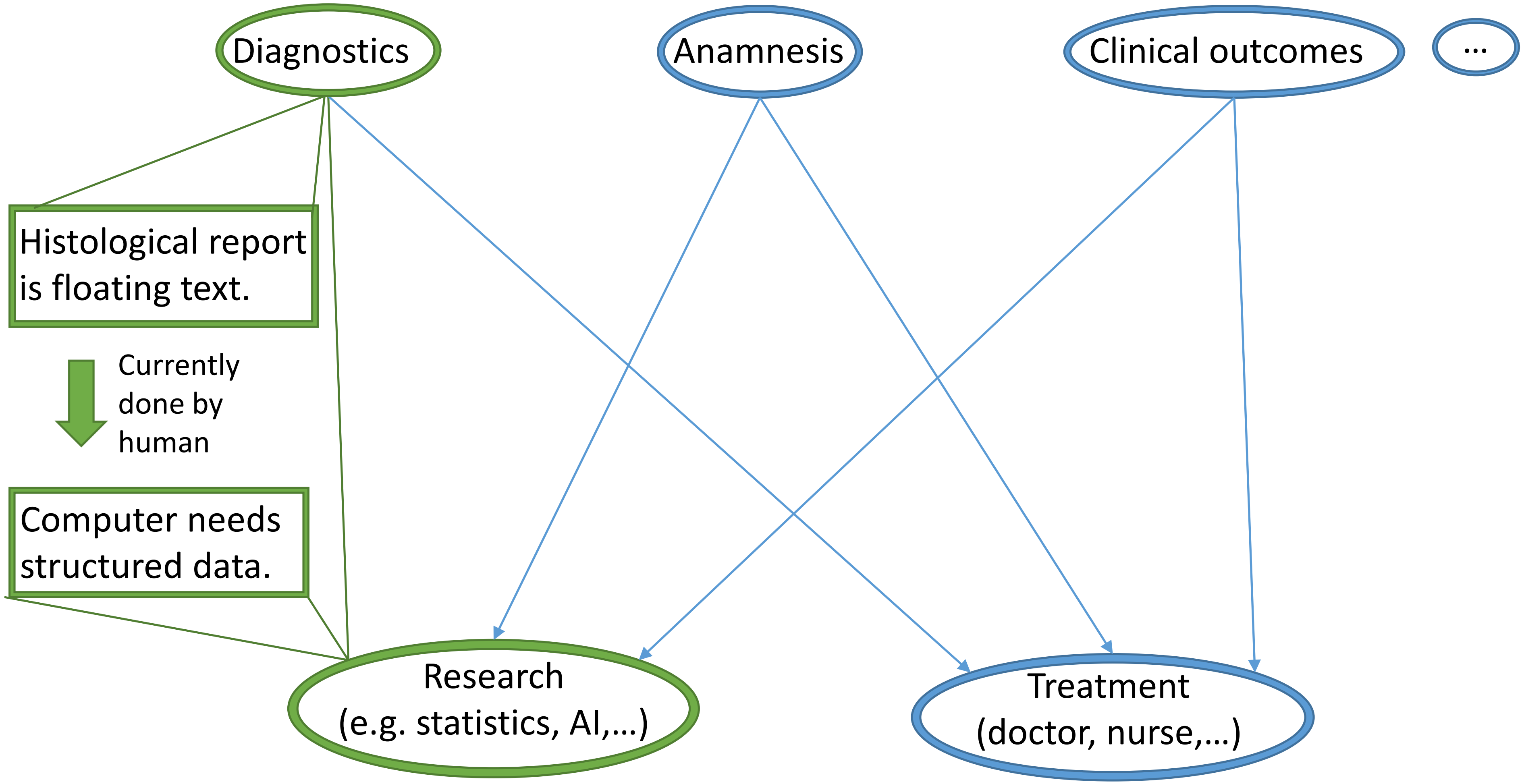
Comparing Link Grammars and Dependency Grammars for parsing German histological reports

Julian Dörenberg[1, 2]
julian.doerenberg@rwth-aachen.de, jdoerenberg@ukaachen.de

[1] RWTH Aachen University
[2] University Hospital RWTH Aachen

Secondary usage of data



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.

(…)“

CENTRALIZED BIOMATERIAL BANK RWTH AACHEN UNIVERSITY

Problem and example sentence

UNIKLINIK RWTH AACHEN

<u>Human reads:</u>

„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)

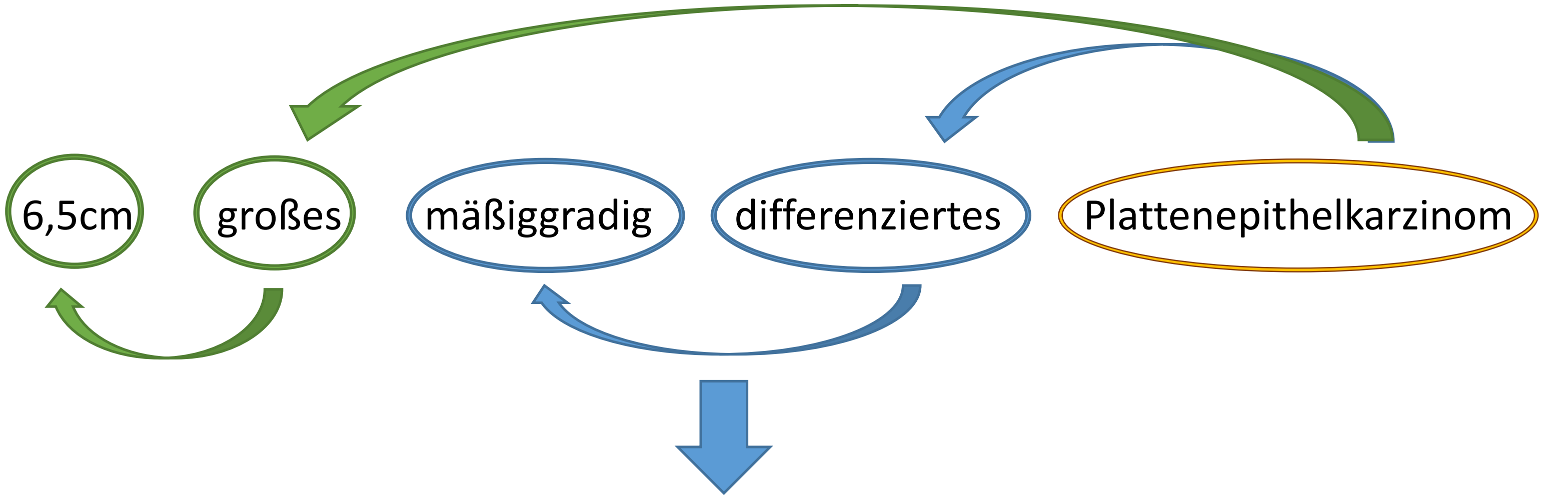
<u>Computer reads:</u>

„Word Word Word Word Word Word Word Word Word.“

-> No structured information

Solution

Relation graph:



Patient ID	Type of Carcinoma	Size	Degree of differenciation
12345	Plattenepithelkarzinom	6,5cm	mäßiggradig

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

Link Grammars [1, 2]

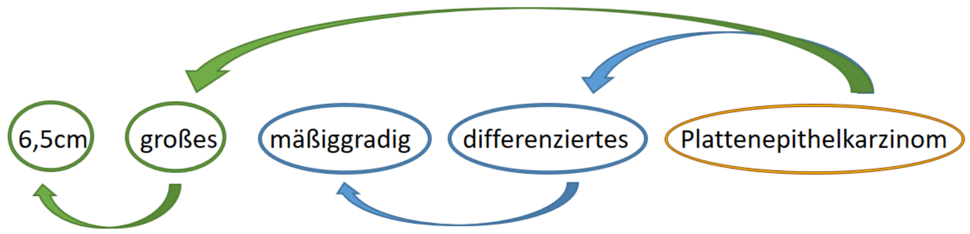
- Grammar is dictionary mapping words to grammatical contexts

Karzinom -> d_{11} V d_{12} V d_{13} …
groß -> d_{21} V d_{22} V d_{23} …
…

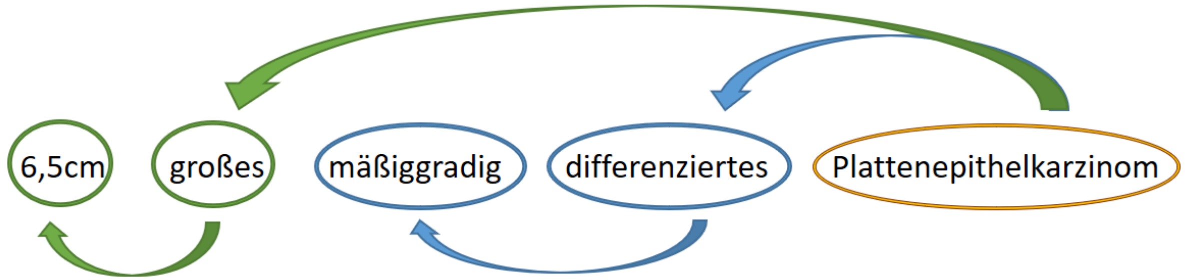
- Parsing is satisfiability checking while complying to meta-rules rules
  - Input: Sentence and grammar
  - Output: Relation graph
  - Idea: Recursively parse regions
    - Base case: Add single relation
    - Backpropagation: Make sure to comply with four rules
  - Runtime: Exponential, but can be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

1) Planarity of relation graphs
2) Connectivity of relation graphs
3) Ordering in disjuncts is maintained
4) Two word are connected by max. one link

- 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 avilable
  - 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)

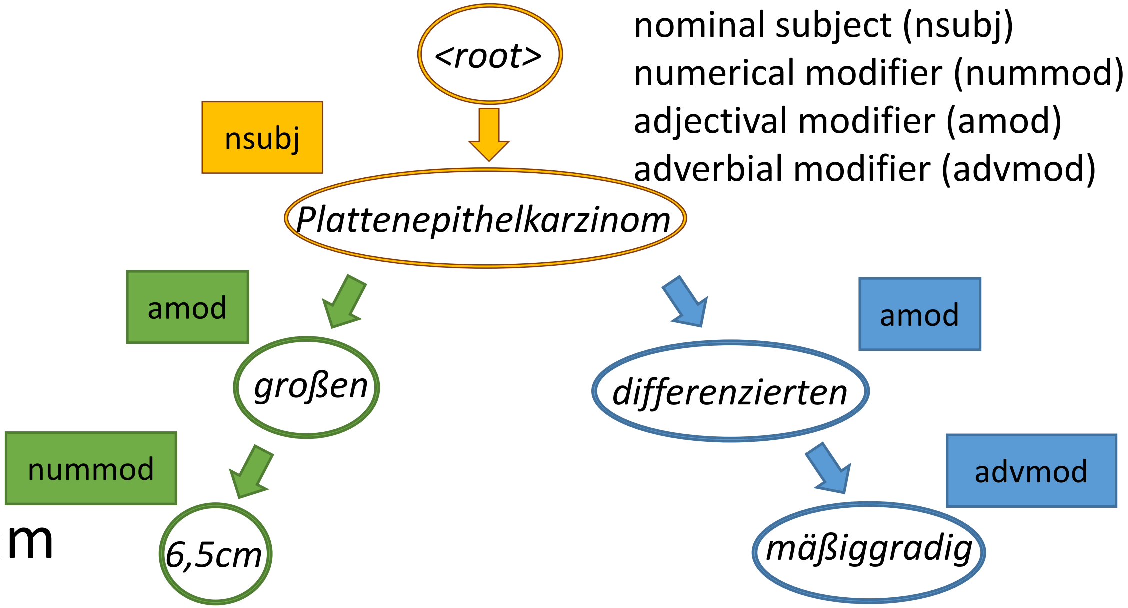
Proof-of-Concept ✓

Dependency Grammars (DGs) [3]

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

Grammatical role of word:
nominal subject (nsubj)
numerical modifier (nummod)
adjectival modifier (amod)
adverbial modifier (advmod)

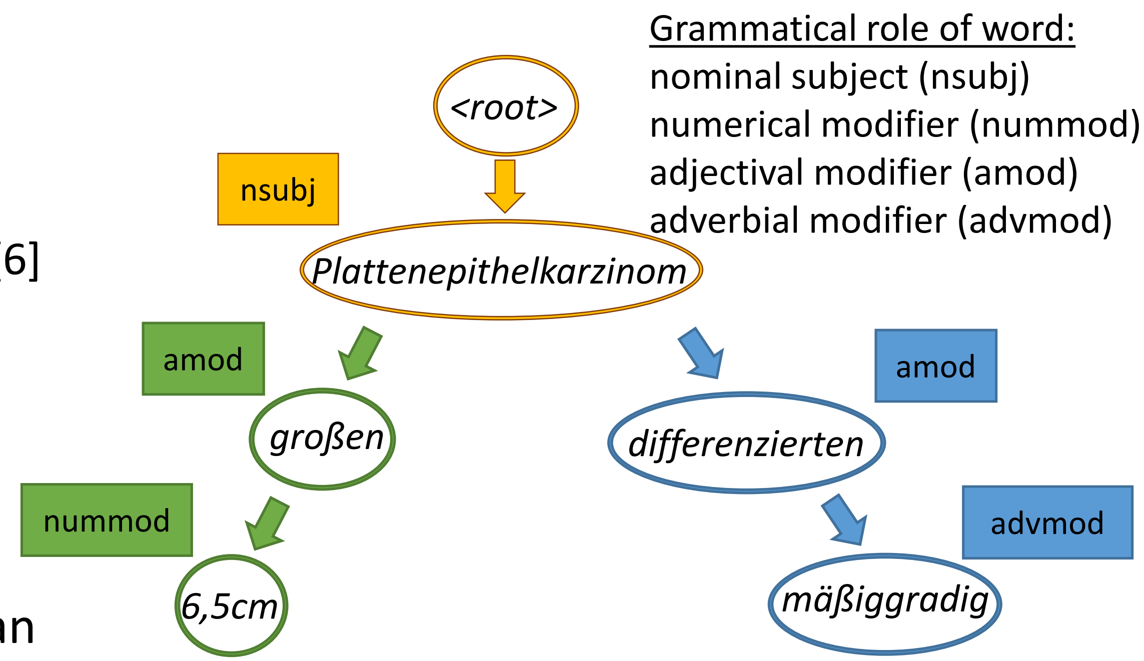
- Parsing is applying RNN operations
  - Input: Embedded words
  - Output: Relation graph tree
  - Idea: RNNs
  - Runtime: Polynomial



- Traning 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

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

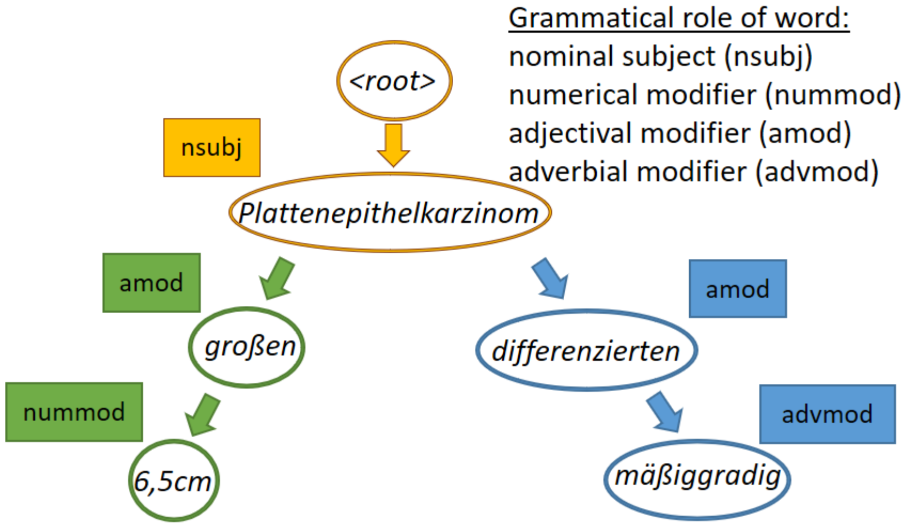
- Evaluation data:
  - 200 sentences from breast biopsy reports
  - Written by two senior pathologists and annotated by me

* 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



(großen, Plattenepithelkarzinom)
(6,5cm, großen, Plattenepithelkarzinom)
(großen, Plattenpithelkarzinom, differenzierten, mäßiggradig)

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

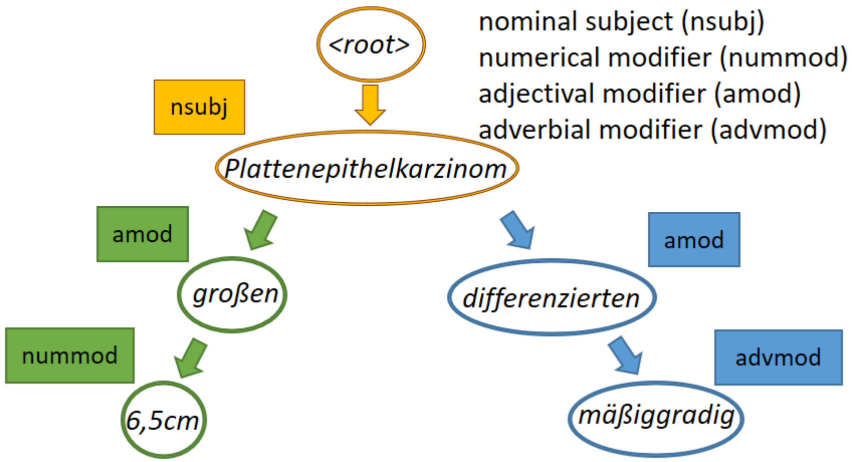
Evaluation of the DG parser

Evaluation data: 200 sentences randomly selected from the breast biopsy reports
-> Localization sentence (local.) do not have unambiguous parsing

Metric	Dozat and Manning [5] All (200 sents)	Dozat and Mannning [5] no local. (165 sents)
Unlabelled Attachment Score (UAS)	0.94	0.96
Labelled Accuracy Score (LA)	0.92	0.95
Labelled Attachment Score (LAS)	0.9	0.93
Proportion of correctly extracted 2-ary relations	0.95	0.97
Proportion of correctly extracted 3-ary relations	0.91	0.93
Proportion of correctly extracted 4-ary relations	0.88	0.89

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



Grammatical role of word:
nominal subject (nsubj)
numerical modifier (nummod)
adjectival modifier (amod)
adverbial modifier (advmod)

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 mor detailed analysis

$\Rightarrow$ 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

Number of HCCs	Fibrosis	Vascular invasion	Tumor diameter	Inflamm.	Inflamm. degree	Distance to resection area	Desmet stage	Steatosis	Cirrhosis
1			1,4cm			1mm			TRUE
1		TRUE	5,5cm			0,3cm			
1	TRUE		4,2cm	TRUE		0,3cm			FALSE
1	TRUE	TRUE	8,5cm	TRUE			3		
1			16cm			0,1cm			
1	TRUE	TRUE	4,2cm	TRUE		1,5mm		TRUE	FALSE
1									
1		FALSE	9,5cm			1cm			
1		FALSE	8,5cm	TRUE				TRUE	
1	TRUE		3,6cm			0,2cm	1-2		

Correct

Wrong

Not given in report

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

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

$\Rightarrow$ DG-annotated training data from histological reports required for improvement of parsing performance

References

[1] Sandra Kübler. 1998. Learning a Lexicalized Grammar for German. New Methods in Language Processing and Computational Natural Language Learning, p. 11-18.

[2] Daniel Sleator, Davy Temperley. 1991. Parsing English with a Link Grammar

[3] Marie-Catherine de Marneffe, Christopher D. Manning, Nivre Joakim, Daniel Zeman. Universal Dependencies. 2017. Computational Linguistics.

[4] https://pypi.org/project/supar/ and https://github.com/yzhangcs/parser

[5] E. Kara, T. Zeen, A. Gabryszak, K. Budde, D. Schmidt, and R. Roller. A Domain-adapted Dependency Parser for German Clinical Text. en. In A Domain-adapted Dependency Parser for German Clinical Text. Verlag der Osterreichischen Akademie der Wissenschaften, 2018.

[6] Timothy Dozat, Christopher D. Manning. Deep Biaffine Attention for Neural Dependency Parsing. 2017. Conference paper at ICLR.

[7] Carlos Gómez-Rodríguez, Iago Alonso-Alonso, David Vilares. How important is syntactic parsing accuracy? An empirical evaluation on rule-based sentiment analysis. 2017. Artificial Intelligence Review.