Computational Narratology: Extracting Tense Clusters from Narrative Texts

Thomas Bögel, Jannik Strötgen, Michael Gertz
Database Systems Research Group, Heidelberg University, Im Neuenheimer Feld 348, 69120 Heidelberg, Germany

Context of This Work

Computational Narratology
- context: Digital Humanities
- facilitate annotations from literary scientists
- support hypotheses [1]
- methods: Natural Language Processing

Temporal Phenomena in Literary Texts

Temporal expressions
- explicit temporal information
- can be extracted automatically (HeidelTime [2])
- less frequent in literary text

Tense information
- tenses: information about temporal structures
- shifts in tenses indicate order phenomena (e.g., prolepsis)

Task: robust annotation of tenses in narrative texts

Temporal Phenomena in Narrative Texts

Prior work
- laborious manual annotations
- automatic systems focus on English
- no existing system for German tense annotation

Data set
- German narrative texts (20th century)
- annotations by literary scientists
- tagset: narratological aspects

NLP Pipeline Architecture [3]

Machine learning interface
- feature extraction and machine learning
- interchangeable algorithms
- goal: predict annotations automatically

Feedback loop on predicted annotations
- manual corrections
- improvement of future predictions (ML)

Use Case: Tense Annotations

Extraction of temporal clusters
- temporal cluster: all tokens governed by the same verb
- exploitation of tense markers (e.g., auxiliaries)
- morphological features & heuristics
- heuristic for sentences with unknown tense
- evaluation: comparison to manual annotations
- high inter-annotator agreement (κ > 0.8)

Ongoing work
- machine learning based system for additional annotations, e.g., narrative levels
- hybrid, self-improving system: heuristics + machine learning

Evaluation Results

<table>
<thead>
<tr>
<th>tense</th>
<th>correctly tagged verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>93.10</td>
</tr>
<tr>
<td>preterite</td>
<td>95.73</td>
</tr>
<tr>
<td>perfect</td>
<td>96.43</td>
</tr>
<tr>
<td>pluperfect</td>
<td>84.71</td>
</tr>
<tr>
<td>future</td>
<td>90.00</td>
</tr>
</tbody>
</table>

⇒ reliable and robust prediction of tense clusters

References


Contact Information:
Thomas Bögel
boegel@uni-hd.de
http://dbs.ifi.uni-heidelberg.de/