Annotation Quality Checking and Its Implications for Design of a Treebank
(in Building the Prague Czech-English Dependency Treebank)

Marie Mikulová and Jan Štěpánek
Charles University in Prague
ÚFAL
Prague Czech-English Dependency Treebank

- Deep syntactic (tectogrammatical) parallel treebank
- Similar to Prague Dependency Treebank 2.0
  - Stand-off annotation
  - 4 layers (word-form, morphological, analytical, tectogrammatical) – differences
- Wall Street Journal part of the Penn Treebank (49,000 sentences)
But the strategy isn't helping much this time.

Tato strategie však tentokrát příliš nepomáhá.
Annotation Procedure

- Tectogrammatical layer only
  - 39 attributes (8.42 per node in PDT 2.0)
  - pre-built tree as an input
- Division into several phases
- Periodic measurement of inter-annotator agreement
- Periodic checking of correctness of the annotation
Annotation Quality Checking

Usual approach:

Annotator 1 → Annotator 2 → Annotator 3

9.2 sentences per hour
5 years at a half-time job
€: 3 x 5 = 15

Too slow and too expensive :-(
Annotation Quality Checking (2)

PDT 2.0 approach:

- Checking of finished data.
- No parallel data at all.
Annotation Quality Checking (3)

PCEDT approach:

- Each annotator checks his/her own data.
- Part of the data parallel.
Checking Procedures

- Invariants, impossible or necessary combinations of the nodes and their attributes
- Source:
  - annotation rules
  - annotators' feedback
  - generalization of the output of an automatic checking procedure: searching for the same surface coverage with different annotation
Checking Procedures (2)

- Implemented in TrEd (based on Perl)
- Output table columns:
  - procedure name
  - type of violation
  - last column: position
- Only accurate procedures (exceptions)
- 50 procedures, 103 possible violations
- 5 categories
Checking Procedures – Attribute

- Only a single attribute is tested, the structure is ignored.
  - Currently, only t_lemma (no other non-structural attribute being annotated)

- Example:
  - Reasons are given for every change in pre-generated tectogrammatical lemma.
Checking Procedures – Structure

- Relation between the governing and dependant node and their attributes
- Examples:
  - The root's functor must be PRED, DENOM, PARTL, or VOCAT.
  - PRED and DENOM are possible only for a root.
  - The adnominal attribute (RSTR) can never depend on a verb.
  - Every negated verb has a #Neg child.
  - #EmpVerb and #EmpNoun are never leaves.
Checking Procedures – Coordination

● “Effective” dependencies

● Examples:
  ● Every coordination has at least two members.
  ● Some functors cannot be coordinated together (inner participant (argument) only with an argument of the same sort).

Chief executives and presidents had come and gone.
Checking Procedures – Links

- Links from the t-layer to the a-layer

Examples:
- For every a-node representing a word (i.e. not punctuation) there must be a link from a t-tree.
- The same a-node can be linked as auxiliary to several t-nodes only if the t-nodes are coordinated, or they or their parents have the same t-lemma, or...
- No links to prepositions from DENOM and VOCAT.
Checking Procedures – Valency

● Each verb and deverbative noun is assigned a valency frame.
● Obligatory modifications omitted on the surface must be added to the t-tree.
● Examples:
  ● Valency frame is assigned where required.
  ● No obligatory modification is missing, no actant is superfluous.
  ● “Copied” node has the same valency frame as its original.
Correction Workflow

Data → Checking procedures → List of violating positions → Empty

Correction:
- Each sentence mentioned just once
- No empty entries
Impact on the Treebank Design

- Checking procedures
  - Find errors
  - Reveal vague annotation rules
  - Appreciation of the annotators
Evaluation of Annotators

- Average error rate per sentence for each annotator
- Ranks remain the same in long-term monitoring

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Errors / Sentences</th>
<th>Errors per Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ma</td>
<td>3 271 / 6 026</td>
<td>0.54</td>
</tr>
<tr>
<td>al</td>
<td>1 214 / 3 213</td>
<td>0.38</td>
</tr>
<tr>
<td>iv</td>
<td>2 648 / 8 125</td>
<td>0.33</td>
</tr>
<tr>
<td>ji</td>
<td>301 / 1 064</td>
<td>0.28</td>
</tr>
<tr>
<td>mi</td>
<td>430 / 1 786</td>
<td>0.24</td>
</tr>
<tr>
<td>ka</td>
<td>1 834 / 8 132</td>
<td>0.23</td>
</tr>
<tr>
<td>le</td>
<td>373 / 1 903</td>
<td>0.20</td>
</tr>
<tr>
<td>ol</td>
<td>1 177 / 6 828</td>
<td>0.17</td>
</tr>
<tr>
<td>ALL</td>
<td>12 139 / 39 609</td>
<td>0.31</td>
</tr>
<tr>
<td>ORIG</td>
<td>119 090 / 34 862</td>
<td>3.42</td>
</tr>
</tbody>
</table>
Refining the Annotation Rules

- Example: “Copied” verb has the same valency frame as its original.
  
  *Peter gave Mary flowers and [he gave] Jane sweets.*

- Metaphoric or phraseological usage:
  
  *For a conflict, he does not have enough attention nor [he has] stomach.*

- One meaning split into several valency frames:
  
  *Company A’s stock closed mixed and company B’s [stock closed] down modestly.*
# Most Common Errors

<table>
<thead>
<tr>
<th>Checking Procedure</th>
<th>Occurrences</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>valency003_2_PAT_missing</td>
<td>883</td>
<td>7.27</td>
</tr>
<tr>
<td>links001_6.1_same_aux</td>
<td>700</td>
<td>5.77</td>
</tr>
<tr>
<td>valency003_2_ACT_missing</td>
<td>623</td>
<td>5.13</td>
</tr>
<tr>
<td>links001_1.1_no_tnode</td>
<td>438</td>
<td>3.61</td>
</tr>
<tr>
<td>valency001_1_no_frame</td>
<td>405</td>
<td>3.34</td>
</tr>
<tr>
<td>valency003_4_wrong_aux</td>
<td>387</td>
<td>3.19</td>
</tr>
<tr>
<td>structure016_1_no_neg</td>
<td>378</td>
<td>3.11</td>
</tr>
<tr>
<td>attribute001_1_t-lemma</td>
<td>352</td>
<td>2.90</td>
</tr>
<tr>
<td>structure003_1_fphr_lemma</td>
<td>348</td>
<td>2.87</td>
</tr>
<tr>
<td>valency003_1_invalid_lemma</td>
<td>345</td>
<td>2.84</td>
</tr>
</tbody>
</table>
Thank you.