Assessing the benefits of partial automatic pre-labelling for frame-semantic annotation

Ines Rehbein, Josef Ruppenhofer & Caroline Sporleder

FEAST 2009
Outline

1 Motivation
2 The Data - FrameNet
   • Frame-Semantic Annotation
3 Experimental Setup
   • Annotation Set-Up
   • Data
   • Study design
4 Results
   • Impact of pre-annotation on annotation time
   • Impact of pre-annotation on annotation quality
   • Impact of pre-annotation quality on human annotation
5 Conclusion and Future Work
Outline

1 Motivation
2 The Data - FrameNet
   • Frame-Semantic Annotation
3 Experimental Setup
   • Annotation Set-Up
   • Data
   • Study design
4 Results
   • Impact of pre-annotation on annotation time
   • Impact of pre-annotation on annotation quality
   • Impact of pre-annotation quality on human annotation
5 Conclusion and Future Work
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Motivation

- Linguistic resources with high-quality manual annotations are a backbone of many supervised NLP scenarios
- Manual annotation of linguistic resources is time-consuming and costly
- How can we annotate a large amount of data and still get good quality?

Can partial automatic pre-labelling speed up the annotation process without sacrificing annotation quality?
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Frame Semantics (Fillmore 1976, 1977, ...) 

- **Semantic frames**
  - are **schematic representations of situations** involving various participants, propositions, and other conceptual roles, each of which is called a frame element (FE)
  - The situations include events, states, and relations
  - Some frames also focus on entities/things

- Frames are connected to each other via **frame-to-frame relations** (e.g. Inheritance (is-a), Perspective on, Subframe, Using, ...)
Frame Semantics (Fillmore 1976, 1977, ...)

Example: Self-motion Frame

- Frame Evoking Elements:
  
  \textit{advance.v}, \textit{climb.v}, \textit{crawl.v}, \textit{hike.v}, \textit{hike.n}, \textit{swim.n}, ...

- Core Frame Elements:
  

- Non-core Frame Elements:
  
  \textit{Co-theme}, \textit{Depictive}, \textit{Duration}, \textit{Manner}, \textit{Time}, ...

[Many others \textit{Self-mover}] RUSHED [back \textit{Goal}] [Wednesday morning \textit{Time}]
Frame Semantic Annotation

- Full-text
  - exhaustive annotation of running text with all different frames and roles that occur in the document

- Lexicographic annotation
  - annotation of instances of particular target words used in particular frames
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work

Rehbein, Ruppenhofer & Sporleder
Partial automatic pre-labelling for frame-semantic annotation
Annotation Set-Up

- **Lexicographic annotation** of FrameNet data
- **6 Annotators** (authors + 3 computational linguistics undergraduates with at least 1 year experience in frame-semantic annotation)

**Annotation process**: decorating automatically derived syntactic constituency trees with semantic role labels using Salto (*Burchardt et al., 2006*)

1. **Frame assignment**: choosing the correct frame for a target lemma from a pull down menu
2. **Role assignment**: draw the available frame element links to the appropriate syntactic constituent(s)
Frame Semantic Annotation with Salto (1)
Frame Semantic Annotation with Salto (2)

Rehbein, Ruppenhofer & Sporleder

Partial automatic pre-labelling for frame-semantic annotation
Frame Semantic Annotation with Salto (3)
Frame Semantic Annotation with Salto (4)
Frame Semantic Annotation with Salto (5)

Rehbein, Ruppenhofer & Sporleder

Partial automatic pre-labelling for frame-semantic annotation
Data

- 360 FrameNet sentences (BNC) exemplifying all the senses defined for 6 different lemmas in FrameNet 1.3

<table>
<thead>
<tr>
<th>Instance</th>
<th>Instances</th>
<th>Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>feel</td>
<td>134</td>
<td>6</td>
</tr>
<tr>
<td>follow</td>
<td>113</td>
<td>3</td>
</tr>
<tr>
<td>look</td>
<td>185</td>
<td>4</td>
</tr>
<tr>
<td>rush</td>
<td>168</td>
<td>2</td>
</tr>
<tr>
<td>scream</td>
<td>148</td>
<td>2</td>
</tr>
<tr>
<td>throw</td>
<td>155</td>
<td>2</td>
</tr>
</tbody>
</table>

- 3 random sets of equal size (120 sentences each)

- 3 versions of each set:
  - **No** pre-annotation, **State-of-the-art**, **Enhanced**
Automatic Pre-Annotation of Frame Assignment
Study design

- Assignment of the 6 annotators to 3 groups of 2 (Group I-III)
- Each annotator experiences all 3 annotation conditions (No pre-annotation, State-of-the-art, Enhanced)
- Order of annotation condition varies between Groups I-III

<table>
<thead>
<tr>
<th></th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>Annotators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>E</td>
<td>S</td>
<td>N</td>
<td>5, 6</td>
</tr>
<tr>
<td>Group II</td>
<td>S</td>
<td>N</td>
<td>E</td>
<td>2, 4</td>
</tr>
<tr>
<td>Group III</td>
<td>N</td>
<td>E</td>
<td>S</td>
<td>1, 3</td>
</tr>
</tbody>
</table>

Table: Annotation condition by order and group

- Training sequence to rule out difficulties with unfamiliar frames and frame elements:
  Total of 240 sentences exemplifying all 6 verbs in all their senses
Data Analysis

- Measures:
  - Precision, Recall, F-score for frame assignment against FrameNet gold standard
  - Annotation time for each text segment

- Analysis of Variance (ANOVA)
  - impact of automatic pre-annotation on annotation time
  - impact of automatic pre-annotation on annotation quality (f-score)
Outline

1. Motivation
2. The Data - FrameNet
   - Frame-Semantic Annotation
3. Experimental Setup
   - Annotation Set-Up
   - Data
   - Study design
4. Results
   - Impact of pre-annotation on annotation time
   - Impact of pre-annotation on annotation quality
   - Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Can pre-annotation of frame assignment speed up the annotation process?

- 2-way ANOVA (within-subjects design), crossing the dependent variable (time) with the order of text segments and condition of pre-annotation

No significant influence of pre-annotation on annotation time (but 5 out of 6 annotators were faster on the text segment with Enhanced pre-annotation)
Can pre-annotation of frame assignment speed up the annotation process? (2)

- Order of text segments has significant influence on time requirement: all but 1 annotator needed most time for the text segment given to them first ($p \leq 0.05$)
  → ongoing training effect

Interaction between training effect and pre-annotation might prevent significant effect of pre-annotation on annotation time
Interaction between pre-annotation and time

<table>
<thead>
<tr>
<th>Order</th>
<th>Annot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N E S</td>
<td>1, 3</td>
</tr>
<tr>
<td>S N E</td>
<td>2, 4</td>
</tr>
<tr>
<td>E S N</td>
<td>5, 6</td>
</tr>
</tbody>
</table>

Rehbein, Ruppenhofer & Sporleder
Partial automatic pre-labelling for frame-semantic annotation
Is annotation quality influenced by automatic pre-annotation?

- 2-way ANOVA (within-subjects design), crossing the dependent variable (f-score) with the order of text segments and condition of pre-annotation

Significant effect ($p \leq 0.05$) for impact of pre-annotation on annotation quality

- All annotators achieved higher quality on Enhanced pre-annotated text segments
- 4 out of 6 annotators achieved higher quality on State-of-the-art pre-annotated text segments
How good does pre-annotation need to be to have a positive effect?

- 4 out of 6 annotators achieved higher f-score on State-of-the-art pre-annotated texts → not statistically significant
- State-of-the-art ASRL system is not yet good enough
  - to significantly speed up the annotation process
  - to improve annotation quality
- No evidence that the error-prone pre-annotation decreases annotation quality
How good does pre-annotation need to be to have a positive effect? (2)

- The 2 annotators who showed a decrease in f-score were in the same Group (Group I: E, S, N)
- Benefit from ongoing training, resulting in higher f-scores for the 3rd text segment (N)
- ANOVA for 4 annotators (Groups II,III):
  - all 4 annotators show decrease in annotation quality for N (compared to S)
  - both types of pre-annotation (S, E) increase f-scores for human annotation quality

Impact of pre-annotation on annotation quality is weakly significant ($p \leq 0.1$)
Do annotators make different types of errors on pre-annotated texts?

Figure: F-Scores per frame for human annotators on different levels of pre-annotation and for state-of-the-art ASRL
Outline

1. Motivation
2. The Data - FrameNet
   • Frame-Semantic Annotation
3. Experimental Setup
   • Annotation Set-Up
   • Data
   • Study design
4. Results
   • Impact of pre-annotation on annotation time
   • Impact of pre-annotation on annotation quality
   • Impact of pre-annotation quality on human annotation
5. Conclusion and Future Work
Assessing the benefits of partial automatic pre-annotation
- Automatic pre-annotation has a **positive effect on quality** of human annotation
- Error-prone automatic pre-annotation does **not decrease quality** of human annotation
- **Strong interaction** between order of text segments (→ ongoing training effect) and annotation condition, masking the benefits of automatic pre-annotation

Future work: annotation experiment controlled for order of text segments
Thank You!

Questions?
### Baselines for automatic pre-annotation (Shalmaneser) and enhanced pre-annotation

<table>
<thead>
<tr>
<th>Seg.</th>
<th>Precision</th>
<th>Recall</th>
<th>f-score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shalmaneser</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>(70/112) 62.5</td>
<td>(70/96) 72.9</td>
<td>67.30</td>
</tr>
<tr>
<td>B</td>
<td>(75/113) 66.4</td>
<td>(75/101) 74.3</td>
<td>70.13</td>
</tr>
<tr>
<td>C</td>
<td>(66/113) 58.4</td>
<td>(66/98) 67.3</td>
<td>62.53</td>
</tr>
<tr>
<td><strong>Enhanced Pre-Annotation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>(104/112) 92.9</td>
<td>(104/111) 93.7</td>
<td>93.30</td>
</tr>
<tr>
<td>B</td>
<td>(103/112) 92.0</td>
<td>(103/112) 92.0</td>
<td>92.00</td>
</tr>
<tr>
<td>C</td>
<td>(99/113) 87.6</td>
<td>(99/113) 87.6</td>
<td>87.60</td>
</tr>
</tbody>
</table>

**Table:** Baselines for automatic pre-annotation (Shalmaneser) and enhanced pre-annotation
### Table: Results for frame assignment: precision, recall, f-score (F), time (t) (frame and role assignment), pre-annotation (p): Non, Enhanced, Shalmaneser

<table>
<thead>
<tr>
<th>Annotator</th>
<th>Precision</th>
<th>Recall</th>
<th>F</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94/103</td>
<td>91.3</td>
<td>86.2</td>
<td>88.68</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>99/107</td>
<td>92.5</td>
<td>88.4</td>
<td>90.40</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>105/111</td>
<td>94.6</td>
<td>96.3</td>
<td>95.44</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>93/105</td>
<td>88.6</td>
<td>83.0</td>
<td>85.71</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>86/98</td>
<td>87.8</td>
<td>76.8</td>
<td>81.93</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>98/106</td>
<td>92.5</td>
<td>86.7</td>
<td>89.51</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>95/107</td>
<td>88.8</td>
<td>84.8</td>
<td>86.75</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td>103/110</td>
<td>93.6</td>
<td>92.0</td>
<td>92.79</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>99/113</td>
<td>87.6</td>
<td>87.6</td>
<td>87.60</td>
<td>117</td>
</tr>
<tr>
<td>4</td>
<td>106/111</td>
<td>95.5</td>
<td>94.6</td>
<td>95.05</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>99/108</td>
<td>91.7</td>
<td>87.6</td>
<td>89.60</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>105/112</td>
<td>93.8</td>
<td>92.9</td>
<td>93.35</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>104/110</td>
<td>94.5</td>
<td>92.9</td>
<td>93.69</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>91/103</td>
<td>88.3</td>
<td>80.5</td>
<td>84.22</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>96/100</td>
<td>96.0</td>
<td>85.0</td>
<td>90.17</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>102/106</td>
<td>96.2</td>
<td>91.1</td>
<td>93.58</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>94/105</td>
<td>89.5</td>
<td>83.9</td>
<td>86.61</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>93/100</td>
<td>93.0</td>
<td>82.3</td>
<td>87.32</td>
<td>135</td>
</tr>
</tbody>
</table>
Interaction between pre-annotation and f-score

<table>
<thead>
<tr>
<th>Order</th>
<th>Annot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N E S</td>
<td>1, 3</td>
</tr>
<tr>
<td>S N E</td>
<td>2, 4</td>
</tr>
<tr>
<td>E S N</td>
<td>5, 6</td>
</tr>
</tbody>
</table>
Table: Average f-scores for the 6 annotators

<table>
<thead>
<tr>
<th>Anot1</th>
<th>Anot2</th>
<th>Anot3</th>
<th>Anot4</th>
<th>Anot5</th>
<th>Anot6</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.2</td>
<td>80.1</td>
<td>87.7</td>
<td>89.2</td>
<td>82.5</td>
<td>84.3</td>
</tr>
</tbody>
</table>

Neither pre-annotation nor order of text segments has significant impact on Semantic Role Assignment