German SRL: Corpus Construction and Model Training
Maxim Konca, Andy Lücking, Alexander Mehler
Goethe University Frankfurt, Text Technology Lab

Abstract

We provide a semantic role-annotated resource for training semantic role models for the German language due to a combined translation and alignment process. The gold standard CoNLL-2012 semantic role annotations are translated into German. Semantic role labels are transferred due to alignment models. The resulting dataset is used to train a German semantic role model. With F1-scores around 0.7, the major roles achieve competitive evaluation scores, but avoid limitations of previous approaches.

Need for new resource

The most prominent corpus that is currently used for SRL in the German language, is CoNLL-2009 (Hajič et al., 2009). However, due to a couple of issues (e.g., incomplete mapping of Salsa to PropBank roles – see paper for details), a new resource for German SRL is needed.

Corpora

• The initial dataset is the English CoNLL-2012 corpus.
• In addition to the PropBank roles we use “PRED” to label the semantic role (licensing predicate).
• The original English CoNLL-2012 corpus was translated using the state-of-the-art machine translation system DeepMo (Tiedemann, 2012).
• Automatic annotation of parallel corpora:
  1. Semantic roles with the model of Zhang et al. (2022) (F1: 0.86).
  2. Token alignments provided by the parallel corpora were used to transfer the arguments to the corresponding German tokens.
• The merged dataset has been split into the German training set and the German test set.

Prediction results

dev set

<table>
<thead>
<tr>
<th>Pred.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Sup.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred.</td>
<td>0.79</td>
<td>0.98</td>
<td>0.87</td>
<td>500</td>
</tr>
</tbody>
</table>

test set

<table>
<thead>
<tr>
<th>Pred.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Sup.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred.</td>
<td>0.77</td>
<td>0.81</td>
<td>0.79</td>
<td>460</td>
</tr>
</tbody>
</table>

Model ~ Annot. 1

<table>
<thead>
<tr>
<th>Arg.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Sup.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG1</td>
<td>0.69</td>
<td>0.73</td>
<td>0.71</td>
<td>115</td>
</tr>
</tbody>
</table>

Model ~ Annot. 2

<table>
<thead>
<tr>
<th>Arg.</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1</th>
<th>Sup.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARG1</td>
<td>0.57</td>
<td>0.69</td>
<td>0.63</td>
<td>128</td>
</tr>
</tbody>
</table>

Conclusions

• For comparisons with a model trained on CoNLL-2009 and with X-SRL (Daza and Frank, 2020) see the paper.
• In particular modifiers seem to be difficult.
• The German SRL resource will be published via LDC.

References


