Improving Zero-Shot-Learning for German Particle Verbs by using Training-Space Restrictions and Local Scaling

Maximilian Köper, Sabine Schulte im Walde, Max Kisselew, Sebastian Padó
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart, Germany
{maximilian.koeper,schulte,kisselmx,pado}@ims.uni-stuttgart.de

**Motivation**
- Consider derivational patterns (e.g., \textit{use} $\rightarrow$ \textit{use} + \textit{ful}) as the result of a compositional process, which combines base term and affix
- Learn a mapping function between base verbs (lachen) and particle verbs (anlachen)
- Main contribution: Experiments with various optimizations

**Dataset**
- Experiments on new dataset (German particle verbs) +
- Existing derivational datasets from:
  - DE: Kisselew 2015 and EN: Lazaridou 2013

**Summary**
- General improvement through training space restriction + Local Scaling
- Particle verbs remain challenging!

**Local Scaling**
- Improve Nearest-Neighbor search
  - LS relies on the average distance $\mu$ of $X$ and $Y$ to their $k$ nearest neighbors

**Semantic Vector Space**
- Built on German (20B) and English (9B) COW 2012 corpora
- Use corpus lemmas+POS pairs (schläfen, V, Haus, NN,...)
- Symmetrical window of size 5 (only nouns, verbs, adjectives)
- 400 dimensions (rdb, ppi, frequency smoothing)
- 460000 lemmas for German , 240000 lemmas for English

**Experimental Setup**
- 10-fold cross-validation
- Baseline: predict base as derived term

**Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Particle Verbs (DE)</th>
<th>Kisselew (DE)</th>
<th>Lazaridou (EN)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Default + LS$_{15}$</td>
<td>Default + LS$_{15}$</td>
<td>Default + LS$_{15}$</td>
</tr>
<tr>
<td>Baseline</td>
<td>10.79%</td>
<td>16.08%</td>
<td>15.36%</td>
</tr>
<tr>
<td>AvgAdd</td>
<td>11.82% +1.28%</td>
<td>24.26% +3.14%</td>
<td>24.19% +2.95%</td>
</tr>
<tr>
<td>BestAdd$_1$</td>
<td>10.22% +1.19%</td>
<td>33.91% +3.97%</td>
<td>27.32% +1.87%</td>
</tr>
<tr>
<td>BestAdd$_2$</td>
<td>14.26% +2.24%</td>
<td>38.50% +4.17%</td>
<td>37.06% +1.40%</td>
</tr>
<tr>
<td>BestAdd$_3$</td>
<td>14.44% +1.97%</td>
<td>38.07% +4.61%</td>
<td>38.49% +2.12%</td>
</tr>
</tbody>
</table>

**Restricting the Training Space (BestAdd)**
- Motivation: particles are ambiguous; thus particle+verb derivations often undergo various meaning shifts
- Training on a set of all particle verbs is counterproductive
- Exemplar-inspired strategy: Restrict training data for a given base verb to the $k$ base verbs with the highest cosine similarity

**Evaluation**
- Test instance: \textit{kleben} is most similar to \textit{legen} (thus rely on orange training observation)

**Data available!!**
www.ims.uni-stuttgart.de/data/