Motivation

Problem setting
- Multi-word expressions (MWEs) can be more or less compositional with respect to their components
- Distributional semantic models (DSMs) can approximate compositional properties with semantic closeness
- Ambiguity represents an obstacle for distributional semantic models

Goals
- Improve prediction of compositionality levels
- Factor out ambiguity

We are interested in two types of German multi-word expression:
- Noun-noun compounds (NNCs)
- Particle Verbs (PVs)

We suggest
- Soft clustering as an approximation of different word senses
- Distributional similarity of an MWE and one constituent in the same cluster indicates strong compositionality.

Examples

<table>
<thead>
<tr>
<th>Multi-Word Expressions</th>
<th>Mean Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ahornblatt</em></td>
<td>maple leaf</td>
</tr>
<tr>
<td><em>Brot Salat</em></td>
<td>green salad</td>
</tr>
<tr>
<td><em>Sauermagen</em></td>
<td>bile</td>
</tr>
<tr>
<td><em>Löwenzahn</em></td>
<td>dandelion</td>
</tr>
<tr>
<td><em>Fliegenpilze</em></td>
<td>headstool</td>
</tr>
<tr>
<td><em>Fresenium</em></td>
<td>shoot chopper</td>
</tr>
<tr>
<td><em>aufgehoben</em></td>
<td>illuminate</td>
</tr>
<tr>
<td><em>aufgehoben</em></td>
<td>illuminate</td>
</tr>
<tr>
<td><em>ausgereizt</em></td>
<td>exhaust</td>
</tr>
<tr>
<td><em>erfallen</em></td>
<td>remember/intrude</td>
</tr>
<tr>
<td><em>aufgepfiffen</em></td>
<td>instigate</td>
</tr>
</tbody>
</table>

Examples of German noun-noun compounds and German particle verbs, accompanied human mean ratings on the degrees of compound-component compositionality (scale from 1 to 6).

Models

We use two types of models:
- Standard word space models
- Clustering models, where semantic distance is measured within each cluster

Combination schemes for clusters:
- Highest value among all clusters

Models vary with respect to:
- Number of clusters
- Window size
- POS of co-occurring words
- Combination scheme
- Setting of thresholds

Results

- Higher $\rho$ scores for NNCs than for PVs
- Window models increase their performance with larger context size
- Clustering models perform better for small context sizes in PVs
- PVs and NNCs profit from different combination schemes
- The number of clusters has no big effect on prediction quality
- The improvement with clustering models is stronger for PVs than for NNCs

Conclusions:
- Soft clustering is a good approximation to real sense distinctions in MWEs and their components
- Factoring out ambiguity helps to improve compositional properties assessment
- Different types of MWEs behave differently with respect to ambiguity

Data & Measures

Corpus:
- SdeWaC (v.3, 880 million words) corpus, POS-tagged and lemmatized

Gold Standards:
- GS-NN: 868 German NNCs, 8 annotators.
- GS-PV: 354 PVs, ratings obtained with Amazon Mechanical Turk.

Measures:
- Spearman’s rank order correlation $\rho$
- Cosine similarity