Factoring Ambiguity out of the Prediction of Compositionality for German Multi-Word Expressions

Stefan Bott, Sabine Schulte im Walde
Institute for Natural Language Processing - University of Stuttgart, Germany
{stefan.bott,schulte}@ims.uni-stuttgart.de

Motivation

Problem setting

- Multi-word expressions (MWEs) can be more or less compositional with respect to their components
- Distributional semantic models (DSMs) can approximate compositionality 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

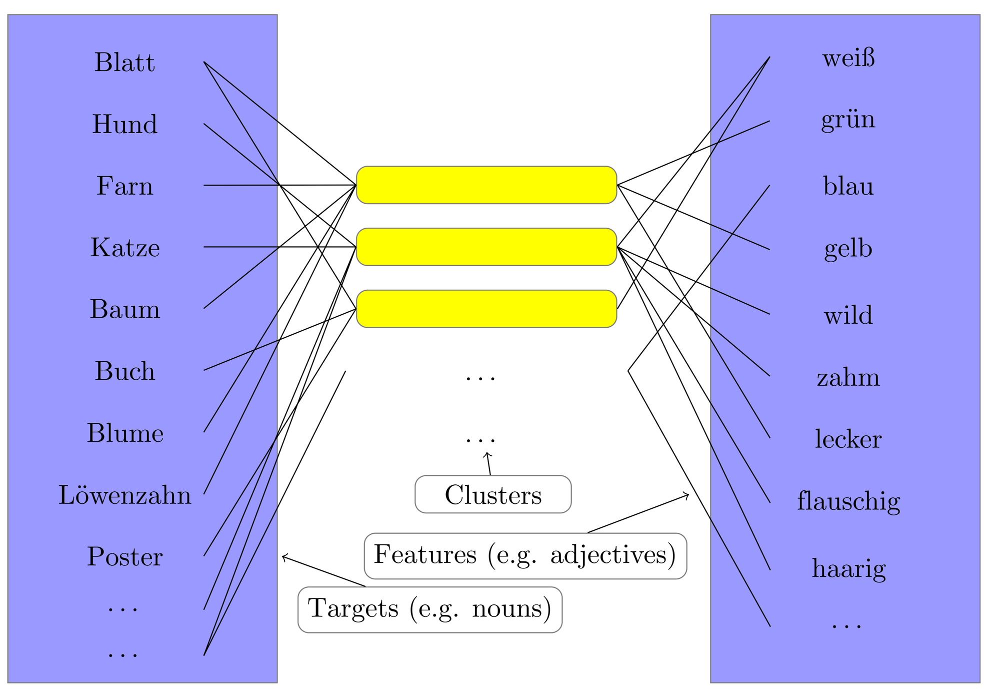
- Soft clustering as an approximation different word senses
- Distributional similarity of an MWE and one constituent in the same cluster indicates strong compositionality.

Softclustering

Latent Semantic Classes (LSC) is a soft clustering algorithm which outputs three probability distributions, across:

- Clusters
- Targets within each cluster
- Predictive feature within each cluster (our features are co-occurring words in local context)

Cluster membership of targets and features can be controlled by thresholds



Data & Measures

Corpus:

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

Gold Standards:

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

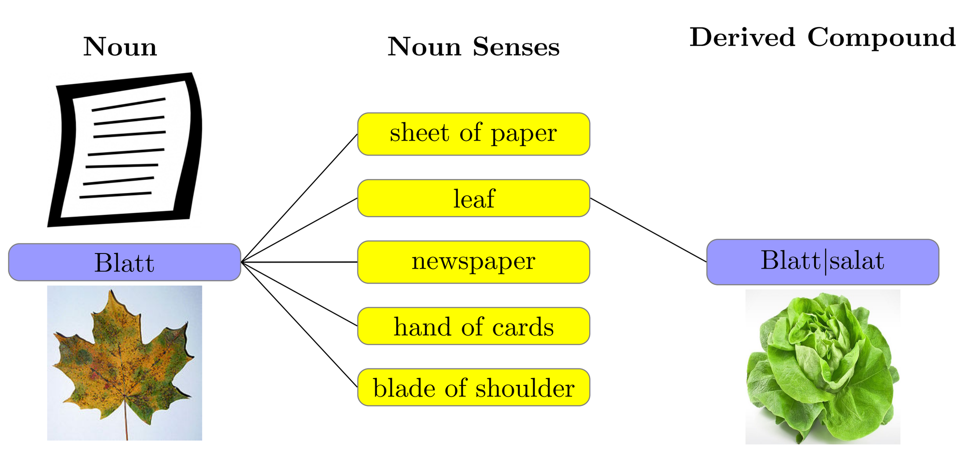
Measures:

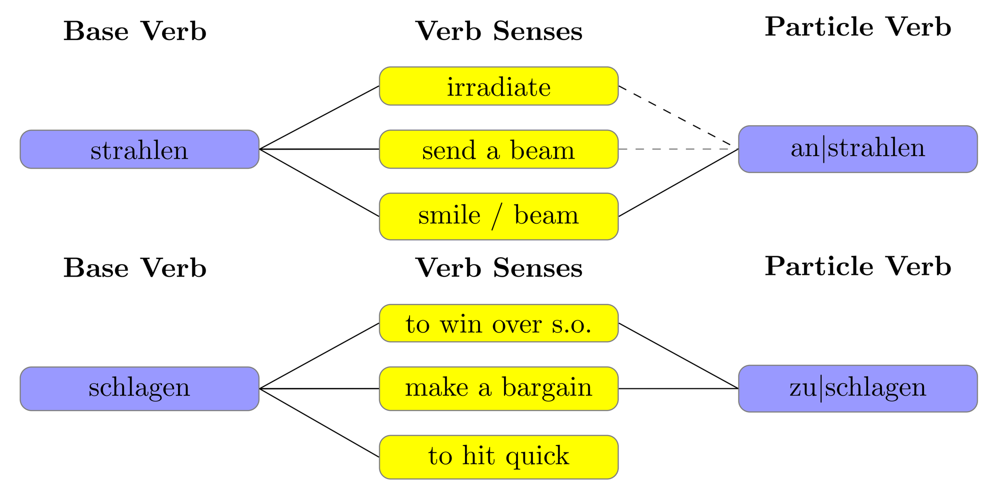
- Spearman's rank order correlation ρ
- Cosine similarity

Examples

Multi-Word Expressions				Mean Ratings Modifier	Mean Ratings Head
Ahorn|blatt	'maple leaf'	maple	leaf	5.64	5.71
Blatt|salat	'green salad'	leaf	salad	3.56	5.68
See|zunge	'sole'	sea	tongue	3.57	3.27
Löwen|zahn	'dandelion'	lion	tooth	2.10	2.23
Fliegen|pilz	'toadstool'	fly/bow tie	mushroom	1.93	6.55
Fleisch|wolf	'meat chopper'	meat	wolf	6.00	1.90
an|leuchten	'illuminate'	an$_{PRT}$	illuminate	–	5.95
auf|horchen	'listen attentively'	auf$_{PRT}$	listen	–	4.55
aus|reizen	'exhaust'	aus$_{PRT}$	provoke	–	3.62
ein|fallen	'remember/invade'	ein$_{PRT}$	fall	–	2.54
an|stiften	'instigate'	an$_{PRT}$	create	–	1.80

Examples of German noun-noun compounds and German particle verbs, accompanied human mean ratings on the degrees of compound-constituent 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

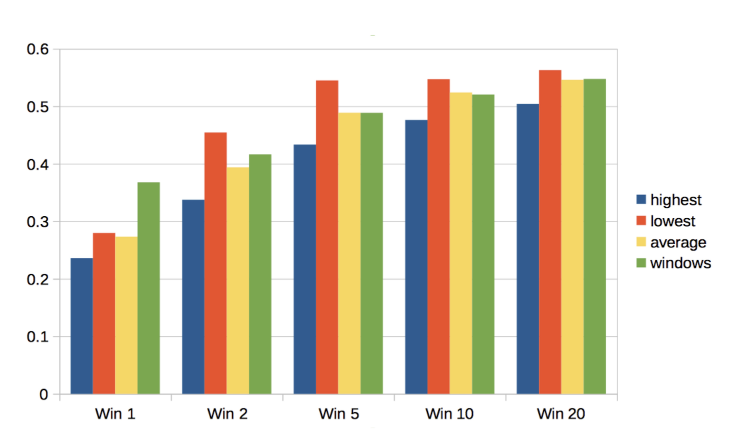
Combination schemes for clusters:

- Highest value among all clusters
- Lowest value among all clusters
- Average value

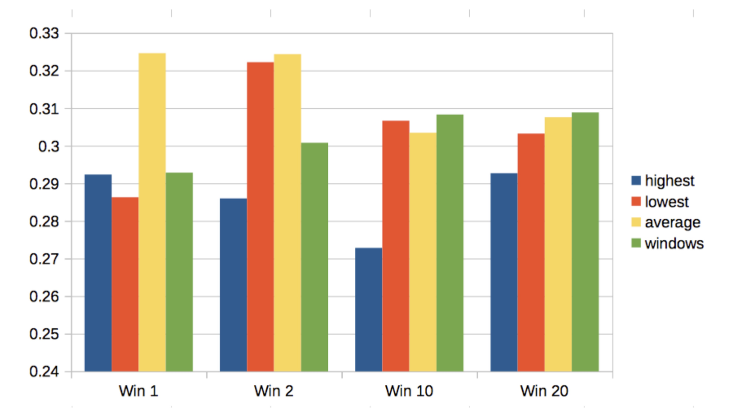
Models vary with respect to:

- Number of clusters
- Window size
- POS of co-occurring words
- Combination scheme
- Setting of thresholds

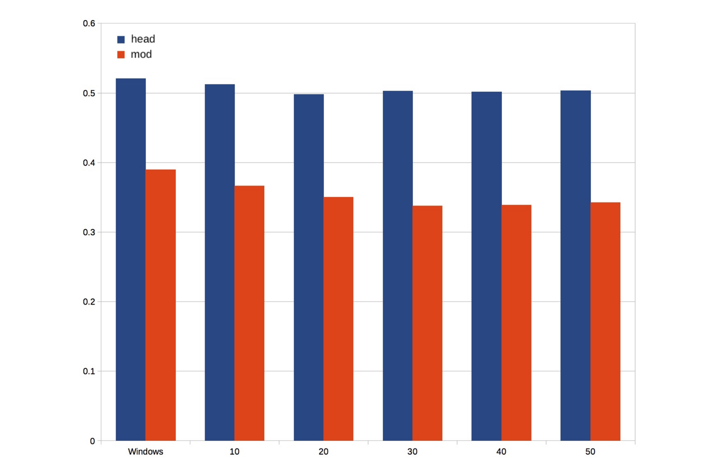
Results



Results (in ρ values) for different window sizes for the NC-head gold standard



Results (in ρ values) for different window sizes for the PV gold standard



Results for different numbers of clusters for the NC gold standard (heads vs. modifiers)

- Higher ρ 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 compositionality assessment
- Different types of MWEs behave differently with respect to ambiguity