Moritz Wittmann[1], Maximilian Köper[2], Sabine Schulte im Walde[2]

[1]iteratec GmbH, Stuttgart, Germany

[2]Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart

Moritz.Wittmann@iteratec.de,
{koepermn, schulte}@ims.uni-stuttgart.de

Exploring Soft-Clustering for German (Particle) Verbs across Frequency Ranges

Goals

We provide an extensive clustering setup and focus on synonymy as a task-independent goal in semantic classification, in order to explore the role of verb frequency ranges across various numbers of clusters. We demonstrate that (1) **low-frequency German verbs are clustered significantly worse than mid- or high-frequency German verbs**, and that (2) **German complex verbs are in general more difficult to cluster than German base verbs**. While (1) the effect of clustering low-frequency target verbs has been investigated by a restricted number of earlier approaches, (2) might be considered as general knowledge but has –as far as we are aware of– not explicitly been proven before.

Data and Algorithm

- The *DECOW14* corpus was processed by *SMOR*, *MarMoT*, and *MATE*.
- 4,871 base verb types and 3,173 particle verb types were extracted.
- Vector spaces for the verbs were obtained from *word2vec* using the *DECOW14* corpus and window sizes of 3 and 10.
- The *Non-negative Matrix Factorization (NMF)* algorithm was used for clustering the verbs, with a k-Means initialization.

Clustering Parameters

- *Verb set*: We clustered either the base verbs, or the particle verbs, or both base and particle verbs, to explore differences for simplex vs. complex verbs.
- *Frequency ranges*: The verbs were sorted by their corpus frequencies, and then split into three equally sized bins. We clustered only verbs from the same frequency range (LOW, MID, HIGH), or all verbs at the same time.
- *Verb vector spaces*: We applied two vector spaces (window sizes 3/10).
- *Number of clusters*: We used 50, 100, and 250 clusters.
- *Number of iterations*: We let the clustering algorithm perform a maximum of 500 iterations (or less if it converged successfully).

Clustering Evaluation

- In general, clustering efforts are motivated by specific tasks or applications, so it is difficult to provide universal recommendations regarding the optimal setup.
- We assess parameters that are generally important on the meta level:
  - *synonymy* as a task-independent goal in semantic classification
  - *compositionality* of a particle verb regarding its base verb (near-synonymy)

Evaluation: Synonymy

- Assessment: pairs of synonymous verbs in the same clusters
- Gold standard: *Duden* dictionary
- Measures: precision, recall, harmonic f-score

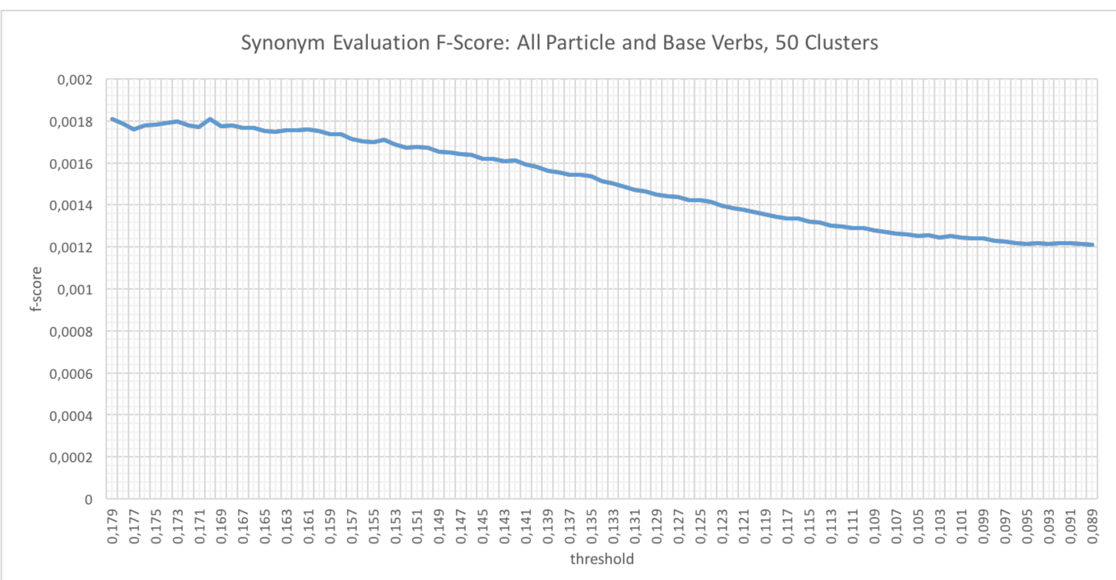
Evaluation: Compositionality

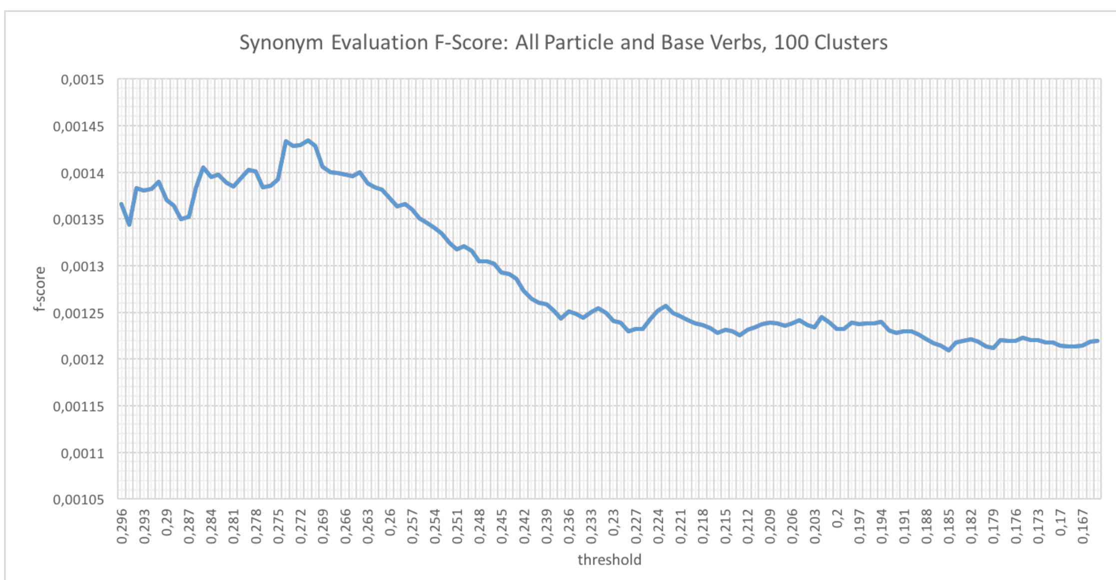
- Assessment: particle and base verbs in the same clusters
  - POINTWISE MUTUAL INFORMATION (PMI): $\log \frac{p(PV,BV)}{p(PV)p(BV)}$, with $p(PV, BV)$ the proportion of clusters containing both the particle verb PV and the base verb BV; $p(PV)$ and $p(BV)$ the proportions of clusters containing the verbs individually
  - COSINE SIMILARITY: cosine between two verbs' average centroid vectors
- Gold Standard: human ratings on particle verb compositionality; contains 400 PVs across 11 particls and 3 frequency bands
- Measure: Spearman's rank-order correlation coefficient ρ

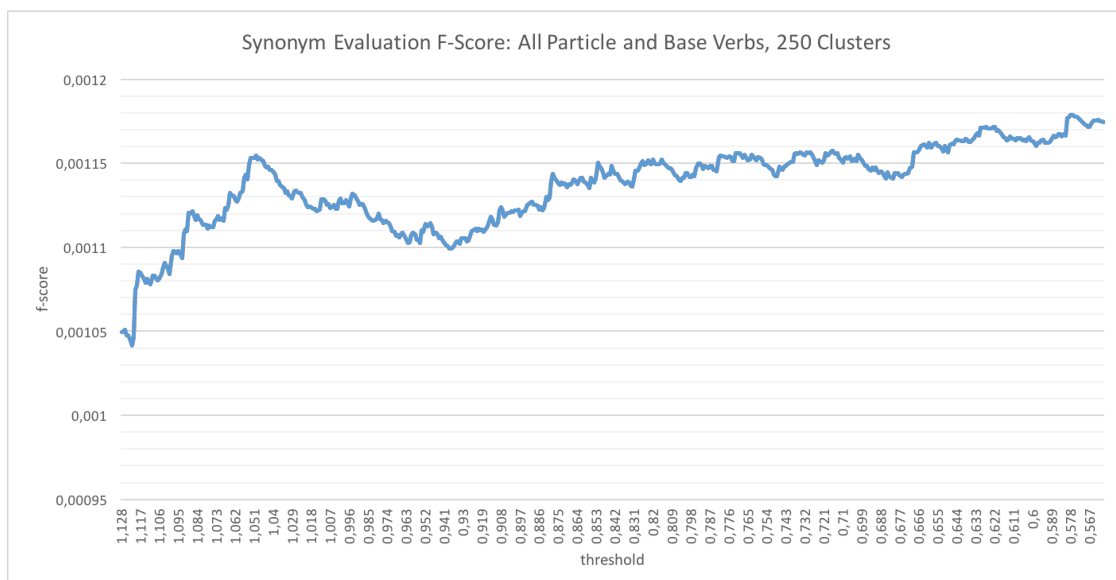
General Parameters of Evaluation

- Techniques for determining cluster membership:
  - THRESHOLD EVALUATION: brute-force search over threshold values: $t_{max} - k \cdot 0.001, k \in \mathbb{N}_0$
  - TOP-N EVALUATION: verbs assigned to the n clusters with highest membership scores, with $1 \leq n \leq \frac{N}{2}$ and N representing the total number of clusters
- Notes:
  - The f-score values for the synonym evaluation are in a very low range because they assess a comparably large number of verb pairs (4,871 base verbs and 3,173 particle verbs) within the cluster analyses.
  - The compositionality evaluation is carried out on a subset of 400 particle verbs for which the gold standard contains compositionality ratings.

Results







Synonymy f-score results for all particle and base verbs and 50/100/250 clusters.

Evaluation: synonymy (threshold)

Frequency	ALL			HIGH			MID			LOW		
Clusters	50	100	250	50	100	250	50	100	250	50	100	250
BVs	**.00640**	.00412	.00370	**.02337**	.01559	.01606	**.00955**	.00480	.00277	**.00212**	.00103	.00090
PVs	.00126	.00076	.00068	.01170	.00602	.00736	.00072	.00025	.00022	.00009	.00004	.00003
BVs+PVs	.00181	.00143	.00118	.01420	.00823	.00925	.00225	.00101	.00084	.00012	.00007	.00004

Evaluation: synonymy (top-n)

Frequency	ALL			HIGH			MID			LOW		
Clusters	50	100	250	50	100	250	50	100	250	50	100	250
BVs	**.01169**	.00736	.00428	**.03006**	.02271	.01999	**.01007**	.00514	.00324	**.00255**	.00144	.00099
PVs	.00217	.00124	.00119	.01335	.00616	.00788	.00088	.00026	.00018	.00004	.00003	.00003
BVs+PVs	.00368	.00351	.00214	.01935	.01206	.00917	.00239	.00152	.00101	.00012	.00007	.00004

Evaluation: compositionality (threshold)

Frequency	ALL			HIGH		
Clusters	50	100	250	50	100	250
BVs+PVs (PMI)	.274***	.183***	.248***	**.468****	.220*	.281**
BVs+PVs (Cos)	**.334*****	.264***	.287***	.439**	.301***	.283**

Evaluation: compositionality (threshold)

Frequency	ALL			HIGH		
Clusters	50	100	250	50	100	250
BVs+PVs (PMI)	.259***	.297***	**.377*****	**.421*****	.378***	.398***
BVs+PVs (Cos)	.197***	.186***	.203***	.311***	.257**	.207*

Results across evaluations and clustering parameters (= p≤0.05, ** = p≤0.01, *** = p≤0.001).*