

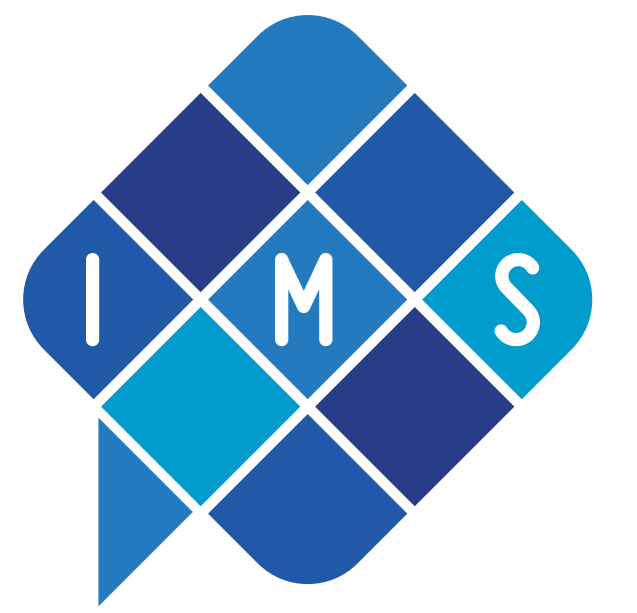
# Applying Multi-Sense Embeddings for German Verbs to Determine Semantic Relatedness and to Detect Non-Literal Language

Maximilian Köper and Sabine Schulte im Walde

Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart, Germany

{maximilian.koeper, schulte}@ims.uni-stuttgart.de



## Motivation

We compare and extend multi-sense embeddings on the challenging class of **German particle verbs**

- Particle + baseverb (e.g. )
- Highly productive
- Often undergo meaning shifts
- Particles are notoriously ambiguous

## Multi-Sense Embeddings

- Learn sense specific vectors:  $\text{aufgeben}_1$  (assign) vs.  $\text{aufgeben}_2$  (surrender)
- Focus on unsupervised methods using **only a corpus** and **no sense inventory**
- Non-parametric methods that learn potentially different numbers of senses per word:

Joint Learning  
**NP-MSSGR**  
 Neelakantan et al.  
(EMNLP 2014)

Joint Learning  
**CHINRESTP**  
 Li and Jurafsky  
(EMNLP 2015)

1Sense→MS  
**CHINWHISP**  
 Pelevina et al.  
(RepL4NLP 2016)

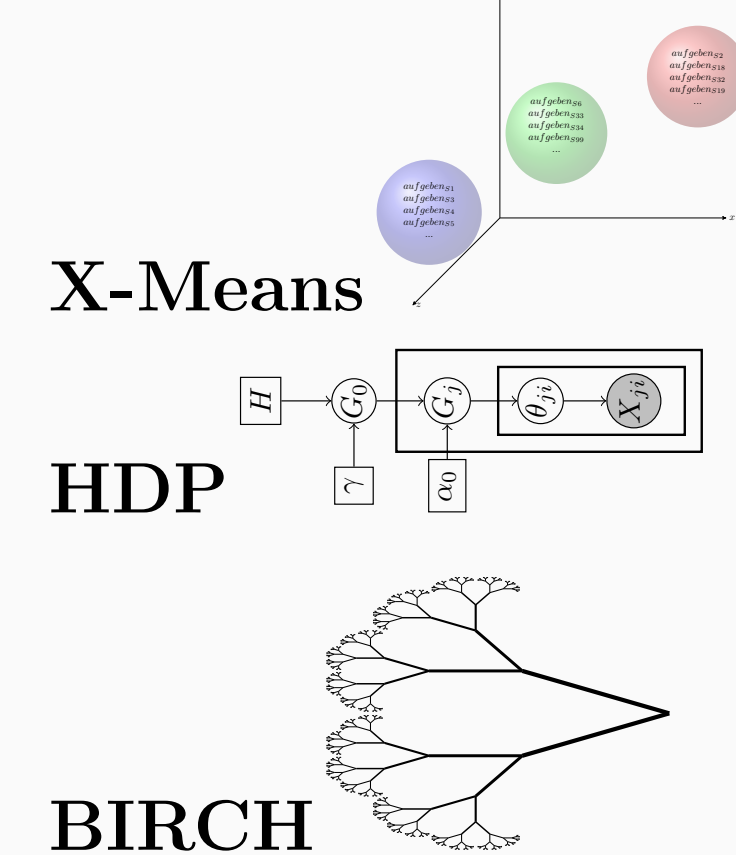
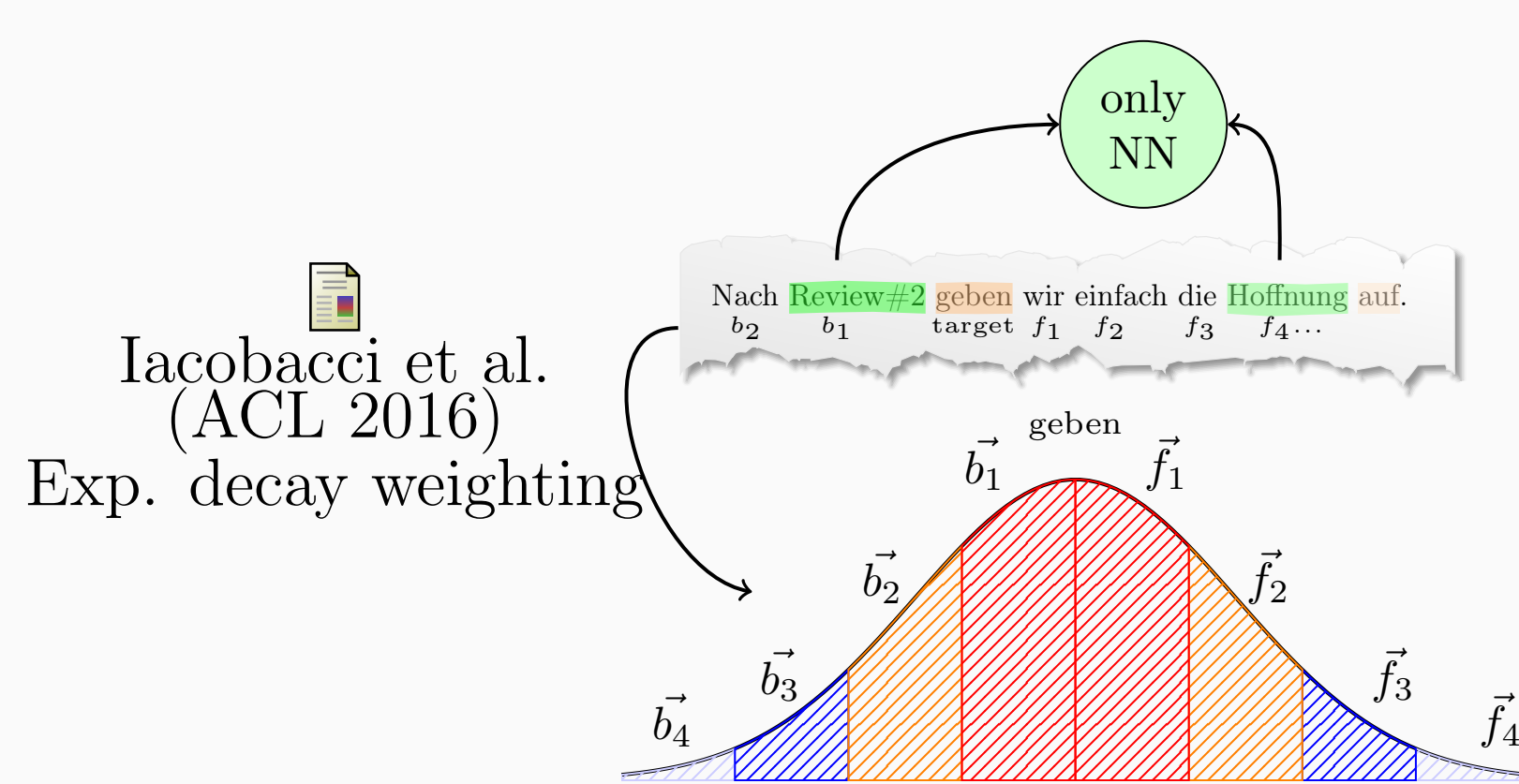
Annotate Corpus with sense information (from clustering) and learn embeddings using *Word2Vec*:  
**HDP, X-MEANS, BIRCH**

## Sense Annotate & Word2Vec

**Note:** Cluster methods are all non-parametric!  
No need to define  $K$  (number of clusters/senses)

- 1) For each of 11 869 verbs, extract all sentences
- 2) Extract features (for each sentence)
- 3) Cluster (each verb separately)
- 4) Create sense-annotated corpus (for all verbs)

Der Anschluss muss von einem Fachmann abgenommen werden ...  
Kim möchte abnehmen und macht deshalb Sport ...  
Nimm doch bitte beim Eintreten ins Haus deine Mütze ab ...  
Er nahm ihr den schweren Koffer ab und trug ihn hinauf ...  
Sie nahm das Telefon erst ab, nachdem es zehnmal geklingelt hatte ...  
Schwätz kein Lällabäbbl, ich nehme dir kein Wort ab! ...  
Jemand soll die Wäsche von der Leine abnehmen ...  
Diese komische Geschichte nehme ich ihm nicht ab.  
Das Bein musste ihm schließlich abgenommen ...  
Wer hat die Prüfung abgenommen?



Der Anschluss muss von einem Fachmann **abgenommen** werden ...  
Kim möchte **abnehmen** und macht deshalb Sport ...  
Nimm doch bitte beim Eintreten ins Haus deine Mütze **ab** ...  
Er **nahm** ihr den schweren Koffer **ab** und trug ihn hinauf ...  
Sie **nahm** das Telefon erst **ab**, nachdem es zehnmal geklingelt hatte ...  
Schwätz kein Lällabäbbl, ich **nehme** dir kein Wort **ab**! ...  
Jemand soll die Wäsche von der Leine **abnehmen** ...  
Diese komische Geschichte **nehme** ich ihm nicht **ab**.  
Das Bein musste ihm schließlich **abgenommen** ...  
Wer hat die Prüfung **abgenommen**?

## Particle Verb Compositionality

We predict the degree of compositionality of German complex verbs and its corresponding base verbs:

- *abnehmen-nehmen* 'take over-take'
- *anfangen-fangen* 'begin-catch'
- Create sense-weighted average vector
- Report Spearman's  $\rho$
- Data contains 400 German particle verbs

Bott et al. (CogALex 2016)

Model	Prediction
NP-MSSGR	.20
ChinRestP	.30
ChinWhisp	<b>.32</b>
HDP	.19
x-Means(NN)	.19
x-Means(w10Exp)	.26
BIRCH(NN)	.28
BIRCH(w10Exp)	.26
Baseline	.26

## Semantic Verb Classification

We target the semantic classification of German complex verbs by applying hard clustering to multi-sense embeddings, rather than using soft clustering.

- Baseline= *Word2Vec* (1Sense)+*C-Means* (soft)
- MultiSense = *K-Means* (hard)
- Report fuzzy **B-Cubed** f-score
- Gold standards
- 3 manual verb class assignments by:

Kliche, Springorum, Roßdeutscher (2011)

- Datasets represent **fuzzy** gold standards

Model	<i>ab</i>	<i>an</i>	<i>auf</i>	<i>all</i>
NP-MSSGR	.12	.18	.15	.05
ChinRestP	.24	<b>.31</b>	.27	<b>.13</b>
ChinWhisp	<b>.26</b>	.30	<b>.28</b>	.11
HDP	.24	.28	.25	.10
x-Means(NN)	.17	.25	.18	.09
x-Means(w10Exp)	.17	.24	.20	.09
BIRCH(NN)	<b>.26</b>	.30	.26	.12
BIRCH(w10Exp)	<b>.26</b>	.32	.25	.12
Baseline	.25	.26	.19	.11

## Detecting Non-Literal Meaning

We explore the token-based classification of German particle verbs into literal vs. non-literal language usage

- Compute cosine between a sense vector and the context words in a sentence
- Add verb-sense-specific token:

Nach Review#2 geben wir einfach die Hoffnung auf. **+aufgebens<sub>1</sub>**

- Perform 10-fold cross-validation
- Multinomial Naive Bayes classifier
- Report results for non-literal class
- Data: 6 436 sentences across 159 verbs:

Köper and Schulte im Walde (NAACL 2016)

Model	P	R	F1
NP-MSSGR	90.1	80.3	84.9
ChinRestP	89.0	79.7	84.1
ChinWhisp	90.1	81.2	85.4
HDP	90.8	80.1	85.1
x-Means(NN)	<b>93.2</b>	<b>83.7</b>	<b>88.2</b>
x-Means(w10Exp)	91.9	81.4	86.3
BIRCH(NN)	91.4	81.6	86.2
BIRCH(w10Exp)	91.1	82.7	86.7
Baseline (K&SiW)	91.1	66.0	76.5

## Hyperparameters&Preprocessing

- DECOW14AX
- Dime: 400
- Non-target words: 1 sense
- Voc.: 200K
- Window: 10
- Other: default
- Max sense: 20
- Negative samp.: 15

**Hard clustering** with **Multi-Sense** Embeddings performs better than **soft clustering** with **Single-Sense** Embeddings!

## Conclusions

- Multi-Sense > Single-Sense
- We need to distinguish between word senses in a distributional semantic model.
- No best multi-sense embeddings across all experiments.
- All embeddings showed high similarity between senses of a word but different behaviour with respect to avg. number of senses per word.