University of Stuttgart
Germany

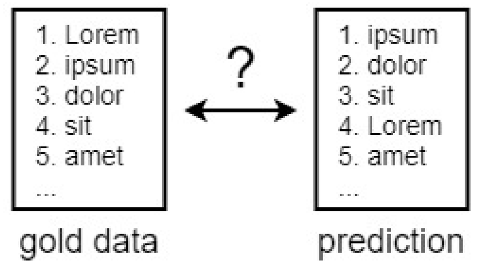
Effects of Pre- and Post-Processing on type-based Embeddings in Lexical Semantic Change Detection

Jens Kaiser Sinan Kurtyigit Serge Kotchourko Dominik Schlechtweg

University of Stuttgart, Institute for Natural Language Processing

Task: Lexical Semantic Change Detection (LSCD)

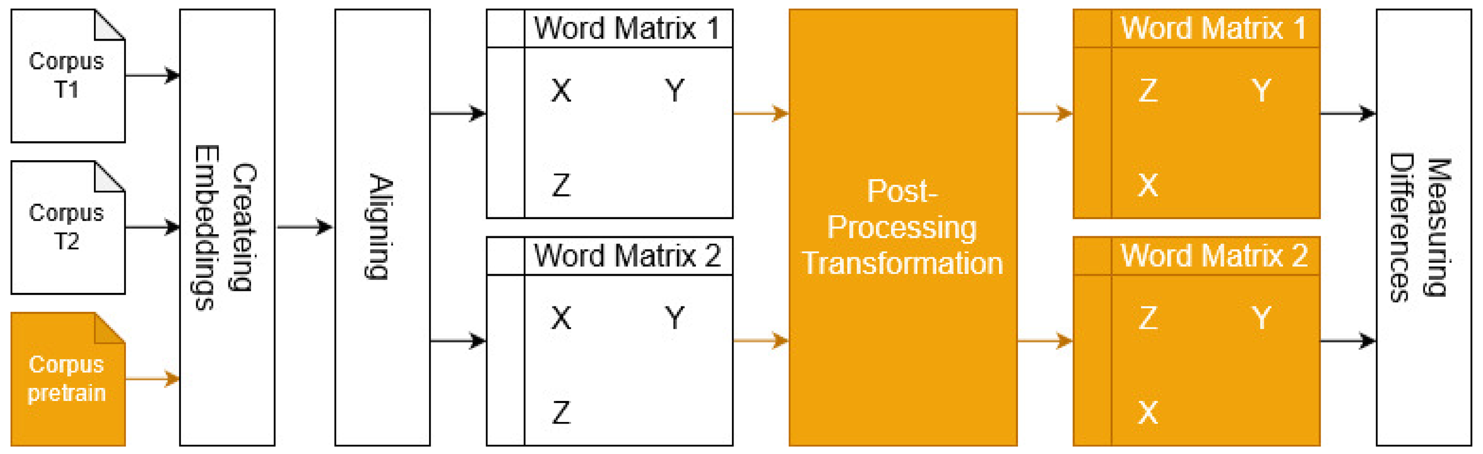
Given two Corpora from different time periods, rank a set of target words according to their degree of semantic change between T1 and T2. Performance on predicted ranking is measured against gold data (human annotation) and quantified by Spearman's rank-order correlation coefficient. [Schlechtweg et al. 2020]



System Description

Traditionally LSCD Models follow the same basic method of using these sub-systems:

1. Creating word representations (using SGNS).
2. Aligning the word representations across corpora.
3. Measuring differences between representations (using Cosine Distance).



- Pre-training enriches the language model with additional general semantic data. We use two corpora MODERN & DIACHRON for pre-training.
- Post-Processing suppressing unwanted information while preserving semantic information.

  - Second Order Transformation (SOT) Word embeddings are transformed by a linear transformation matrix to derive higher or lower orders of similarity. These orders are believed to capture different aspects of language. [Artetxe et al. 2018]

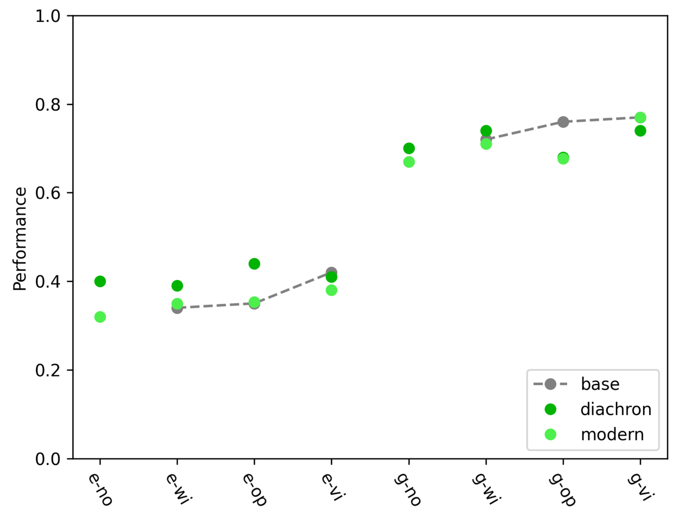
  - Principal Component Removal (PCR) Calculates the principal components of the embedding space and nullifies the top *m* components, which encode word frequency. [Mu et al. 2018]

Research Aim

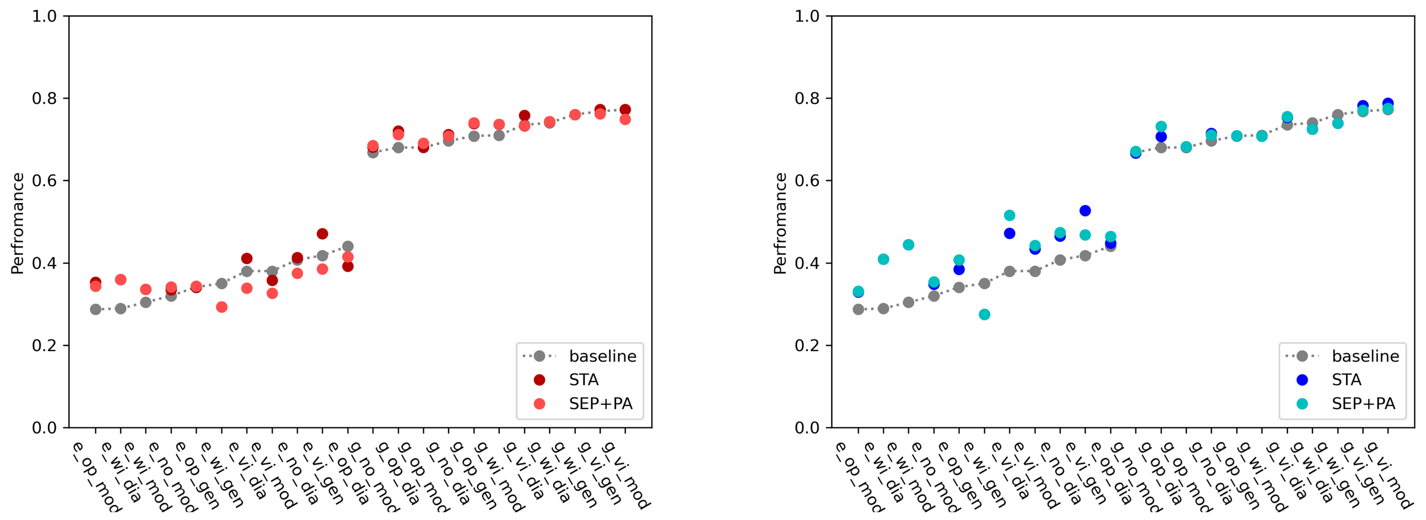
Pre-training and post-processing has successfully been utilized to improve results on semantic similarity and analogy tasks. Does modern pre-training and matrix post-processing improve performance on LSCD?

Improving upon the Baseline

- datasets: SemEval-2020 German (g) and English (e)
- alignment: Vector Initialization (VI), Orthogonal Procrustes (OP), Word Injection (WI) and no alignment (NO)



(a) Pre-training: Performance (Spearman) of the different alignment methods on German and English



(b) Post-Processing: **Left** SOT, **Right** PCR; Performance (Spearman) of the different alignment methods on German and English, with and without (gen) pre-training. "STA" and "SEP+PA" denote different implementations.

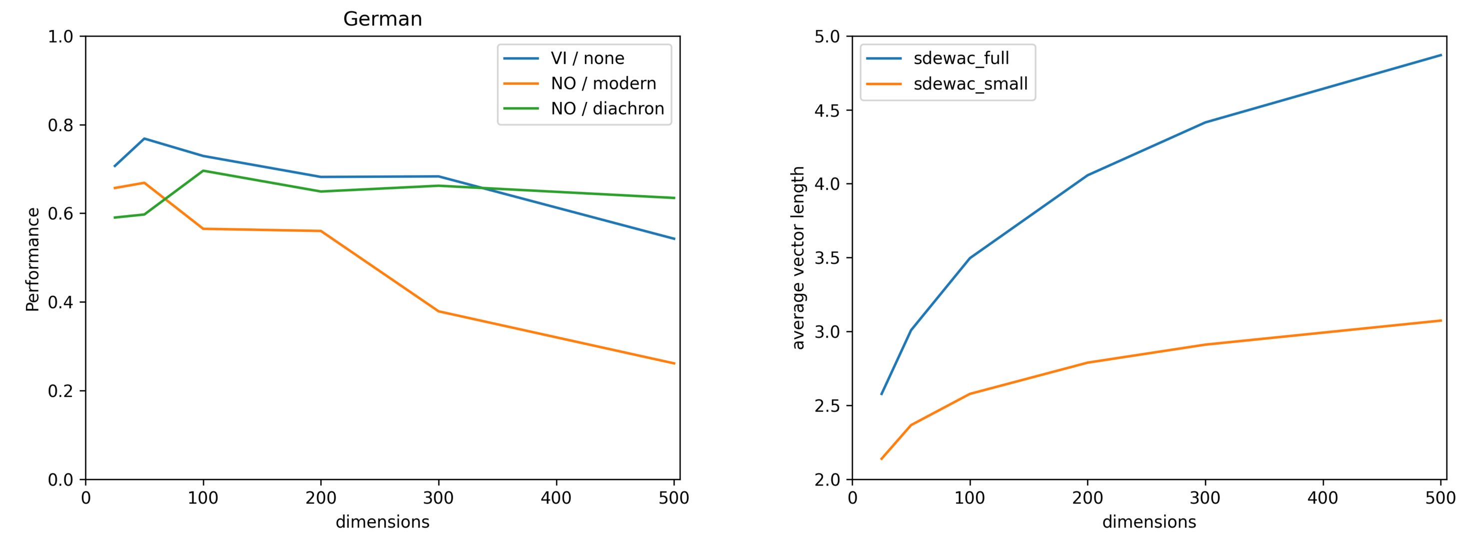
Figure 1a,b: max performance across datasets, alignments method (and pre-training corpora). In most cases improvements, but only at very specific parameter settings.

Parameter fine-tuning on test data is necessary. This poses a challenge to most LSCD tasks as they only provide evaluation data without tuning data available for fine-tuning.

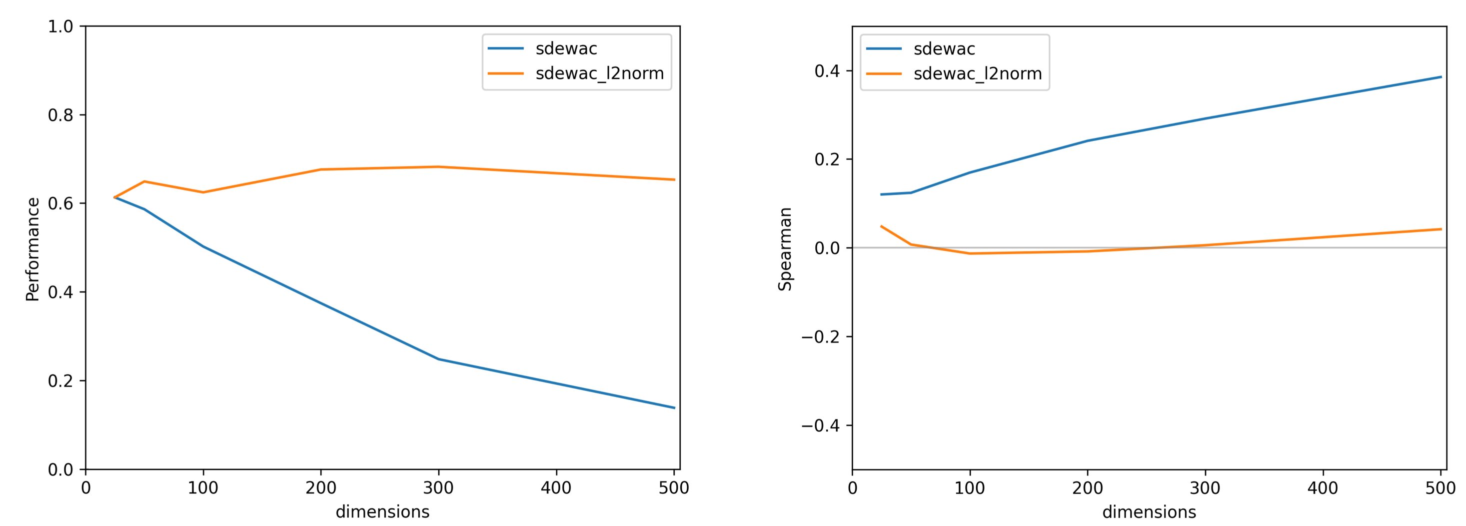
Dimensionality, Frequency Bias, Vector Length

Analyzing the effects of pre-training we observe a connection between performance, frequency bias and vector length at different word vector dimensionality.

- Frequency Bias: Correlation between measured semantic change and word frequency
- Vector Length: Average length of the word vectors created by the language model (SGNS)



(a) **Left:** Comparing performance with different pre-training (none / modern / diachron) across dimensions.
Right: Average vector length for differently sized corpora across dimensions.



(b) **Left:** Performance with and without normalization (l2norm) across dimensions.
Right: Frequency Bias with and without normalization (l2norm) across dimensions.

We observe the following:

- more pre-training data → longer vectors
- more pre-training data → higher Frequency Bias
- more pre-training data → lower performance at high dimensions

Length normalizing word vectors after pre-training removes the Frequency Bias and at the same time avoids a performance drop at higher dimensions. This allows for a broader spectrum of viable parameter settings and the use of large pre-training corpora without introducing Frequency Bias.

References

Artetxe, M., Labaka, G., & Agirre, E. (2018, February). Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of the thirty-second aaai conference on artificial intelligence* (pp. 5012–5019).

Mu, J., & Viswanath, P. (2018). All-but-the-top: Simple and effective postprocessing for word representations. In *International conference on learning representations.* Retrieved from https://openreview.net/forum?id=HkuGJ3kCb

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., & Tahmasebi, N. (2020). SemEval-2020 Task 1: Unsupervised Lexical Semantic Change Detection. In *Proceedings of the 14th international workshop on semantic evaluation.* Barcelona, Spain: Association for Computational Linguistics.