University of Stuttgart
Germany

A Wind of Change

Detecting and Evaluating Lexical Semantic Change across Times and Domains

July 26, 2019

D. Schlechtweg, A. Hätty, M. Del Tredici, S. Schulte im Walde
dominik.schlechtweg@ims.uni-stuttgart.de
Institute for Natural Language Processing, University of Stuttgart, Germany

Motivation

- **evaluation** in research on Lexical Semantic Change Detection (LSCD) is still an unsolved issue (e.g. Cook, Lau, McCarthy, & Baldwin, 2014; Frermann & Lapata, 2016; Lau, Cook, McCarthy, Newman, & Baldwin, 2012; Takamura, Nagata, & Kawasaki, 2017)
- many different modeling approaches coexist
- models are evaluated only superficially, while some of their predictions can be shown to be biased (Dubossarsky, Weinshall, & Grossman, 2017).

→ **we perform the first large-scale evaluation for LSCD**

Evaluation Framework

- evaluation framework and data proposed in Schlechtweg, Schulte im Walde, and Eckmann (2018)
- reduces LSCD to a comparison of word uses in **2 time-specific corpora**

LSC Example

EARLIER

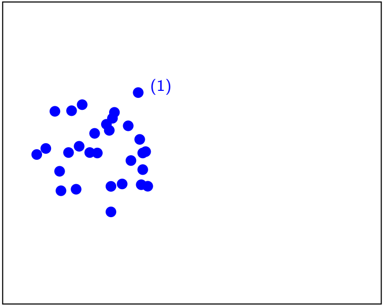
(1) *An schrecklichen Donnerwettern und heftigen Regengüssen fehlt es hier auch nicht.*
‘There is no lack of horrible thunderstorms and heavy rainstorms.’

LATER

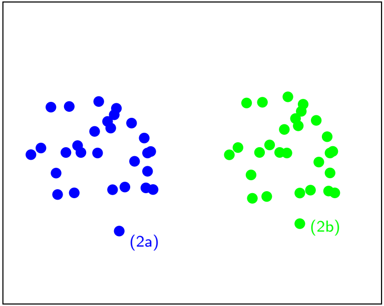
(2) a) *Oder es überschauerte ihn wie ein Donnerwetter mit Platzregen.*
‘Or he was doused like a thunderstorm with a heavy shower.’

b) *Potz Donnerwetter!*
‘Man alive!”

DURel



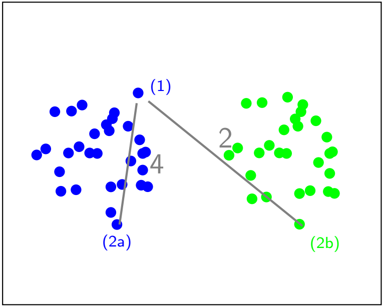
a: EARLIER



b: LATER

Figure 1: 2-dimensional use spaces (semantic constellation) in two time periods with a target word *w* undergoing innovative meaning change. Dots represent uses of *w*. Spatial proximity of two uses means high relatedness.

DURel COMPARE



EARLIER + LATER

From DURel to SURel

- **diachronic LSC detection**: from one time period to another
- **synchronic LSC detection**: from general-language to domain-specific use

Datasets

- **DURel**: rank of 22 target words annotated across time periods
  - a: 1750–1799
  - b: 1850–1899
- **SURel**: rank of 22 target words annotated across domains
  - a: general-language
  - b: domain-specific

Corpora

	Times		Domains	
	DTA18	DTA19	SDEWAC	COOK
size	26,650k	40,323k	109,731k	1,049k

Table 1: Corpora and their sizes.

Task

Given two corpora C_a and C_b,

- rank all target words according to their degree of LSC between C_a and C_b as annotated by human judges;

LSCD Models

- unsupervised
- distributional
- bag-of-words-based
- differ by
  1. **semantic representation type**:
     - semantic vector spaces
     - topic distributions
  2. **alignment methods**
  3. **LSCD measures**

Semantic Representation Type

- **Semantic Vector Spaces**
  - *Count-based Vectors*
    - raw count
    - Positive Pointwise Mutual Information (PPMI)
    - Singular Value Decomposition (SVD)
    - Random Indexing (RI)
  - *Predicted Vectors*
    - Skip-Gram with Negative Sampling (SGNS)
- **Topic Distributions**
  - Sense ChANge (SCAN)

Alignment

- **Count alignment**
  - Column Intersection (CI)
- **RI alignment**
  - Shared Random Vectors (SRV)
- **Embedding alignment**
  - Orthogonal Procrustes (OP)
  - Vector Initialization (VI)
- **Word Injection (WI)**

Measure

- **Similarity Measures**
  - Cosine Distance (CD)
  - Local Neighborhood Distance (LND)
  - Jensen-Shannon Distance (JSD)
- **Dispersion Measures**
  - Frequency Difference (FD)
  - Type Difference (TD)
  - Entropy Difference (HD)

Combination Overview

Sem. Repr.	Alignment					Measure					
	CI	SRV	OP	VI	WI	CD	LND	JSD	FD	TD	HD
count	x				x	x	x			x	x
PPMI	x				x	x	x				
PPMI+SVD			x		x	x	x				
RI		x	x		x	x	x				
SGNS			x	x	x	x	x				
SCAN								x			(x)

Table 2: Combinations of semantic representation, alignment types and measures. (FD has been computed directly from the corpus.)

Example of Model Pipeline

18th century	19th century
1786 magna tempestas, so heißt es Sturm, **Donnerwetter**, Wind, u. s.f. und der Deutsche sagt: es kam ein Wetter, ein rechtes Wetter.	1845 Ich habe Erdstöße gefühlt bei heiterer Luft und frischem Ostwinde, wie bei Regen und **Donnerwetter**.
1794 Als wir zwischen dem 30 sten und 35sten Grade südlicher Breite waren, hatten wir sehr oft **Donnerwetter** mit Regen, Hagel oder Schnee, welcher jedoch sogleich schmolz.	1871 so ließ der alte grämliche Herr manchmal ein gewaltiges **Donnerwetter** los, an welches indessen die Minister schon gewöhnt waren, und aus dem sie sich nichts machten.
1796 Ein paar **Donnerwetter** nebst etwas Regen trugen noch mehr zur Kühle bey	1875 Potz **Donnerwetter**, bin aber ich g'loffen!

Preprocessing

18th century	19th century
1786 heißen:VV Sturm:NN **Donnerwetter:NN** Wind:NN Deutsch:NN sagen:VV kommen:VV Wetter:NN recht:ADJ Wetter:NN	1845 Erdstoß:NN fühlen:VV heiter:ADJ Luft:NN frisch Ostwind:NN Regen:NN **Donnerwetter:NN**
1794 Grad:NN südlich:ADJ Breite:NN **Donnerwetter:NN** Regen:NN Hagel:NN Schnee:NN schmelzen:VV	1871 lassen:VV alt:ADJ grämlich:ADJ Herr:NN gewaltig:ADJ **Donnerwetter:NN** Minister:NN gewöhnen:VV machen:VV
1796 **Donnerwetter:NN** Regen:NN tragen:VV Kühle:NN	1875 **Donnerwetter** laufen:VV

Finding Context (Bags of Words)

18th century	19th century
1786 heißen:VV **Sturm:NN** **Donnerwetter:NN** **Wind:NN** Deutsch:NN sagen:VV kommen:VV Wetter:NN recht:ADJ Wetter:NN	1845 Erdstoß:NN fühlen:VV heiter:ADJ Luft:NN frisch Ostwind:NN **Regen:NN** **Donnerwetter:NN**
1794 Grad:NN südlich:ADJ **Breite:NN** **Donnerwetter:NN** **Regen:NN** Hagel:NN Schnee:NN schmelzen:VV	1871 lassen:VV alt:ADJ grämlich:ADJ Herr:NN **gewaltig:ADJ** **Donnerwetter:NN** **Minister:NN** gewöhnen:VV machen:VV
1796 **Donnerwetter:NN** **Regen:NN** tragen:VV Kühle:NN	1875 **Donnerwetter** **laufen:VV**

Building Semantic Representation

	Sturm:NN	Regen:NN	Minister:NN	...
Donnerwetter:NN$_{18c}$	1	2	0	...
Donnerwetter:NN$_{19c}$	0	1	1	...

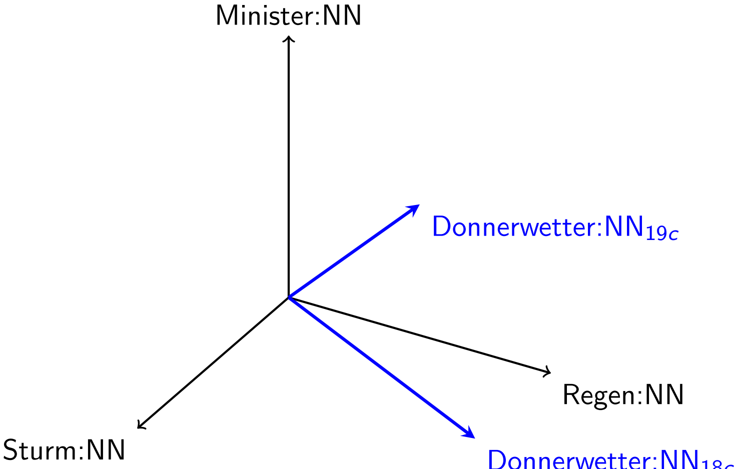
Table 3: Sample table for raw count vectors (we count the number of contexts). Rows contain target words, while columns contain context words. The cells contain the number of co-occurrences between the respective target and context word.

Alignment

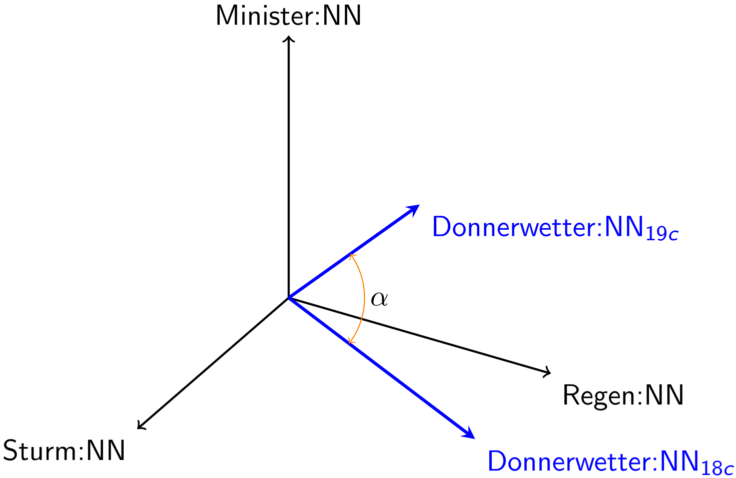
	Sturm:NN	Regen:NN	Minister:NN	...
Donnerwetter:NN_{18c}	**1**	2	0	...
Donnerwetter:NN_{19c}	**0**	1	1	...

Table 4: Sample table for raw count vectors (we count the number of contexts). Rows contain target words, while columns contain context words. The cells contain the number of co-occurrences between the respective target and context word.

Vector Space Interpretation



Cosine Distance



Evaluation Metrics

- **Spearman's rank correlation coefficient** ρ

Best Results

Dataset	Preproc	Win	Space	Parameters	Align	Measure	Spearman m (h, l)
DURel	L_{ALL}	10	SGNS	k=1,t=None	OP	CD	**0.866** (0.914, 0.816)
	L_{ALL}	10	SGNS	k=5,t=None	OP	CD	0.857 (0.891, 0.830)
	L_{ALL}	5	SGNS	k=5,t=0.001	OP	CD	0.835 (0.872, 0.814)
	L_{ALL}	10	SGNS	k=5,t=0.001	OP	CD	0.826 (0.863, 0.768)
	L/P	2	SGNS	k=5,t=None	OP	CD	0.825 (0.826, 0.818)
SURel	L/P	2	SGNS	k=1,t=0.001	OP	CD	**0.851** (0.851, 0.851)
	L/P	2	SGNS	k=5,t=None	OP	CD	0.850 (0.850, 0.850)
	L/P	2	SGNS	k=5,t=0.001	OP	CD	0.834 (0.838, 0.828)
	L/P	2	SGNS	k=5,t=0.001	OP_{-}	CD	0.831 (0.836, 0.817)
	L/P	2	SGNS	k=5,t=0.001	OP	CD	0.829 (0.832, 0.823)

Table 5: Best results of ρ scores (Win=Window Size, Preproc=Preprocessing, Align=Alignment, k=negative sampling, t=subsampling, Spearman m(h,l): mean, highest and lowest results).

Mean Results

Dataset	Representation	best	mean
DURel	raw count	0.639	0.395
	PPMI	0.670	0.489
	SVD	0.728	0.498
	RI	0.601	0.374
	SGNS	**0.866**	**0.502**
	SCAN	0.327	0.156
SURel	raw count	0.599	0.120
	PPMI	0.791	0.500
	SVD	0.639	0.300
	RI	0.622	0.299
	SGNS	**0.851**	**0.520**
	SCAN	0.082	-0.244

Table 6: Best and mean ρ scores across similarity measures (CD, LND, JSD) on semantic representations.

Alignment Results

Dataset	$\mathbf{OP}$	$\mathbf{OP_-}$	$\mathbf{OP_+}$	**WI**	**None**
DURel	0.618	0.557	**0.621**	0.468	0.254
SURel	**0.590**	0.514	0.401	0.492	0.285

Table 7: Mean ρ scores for CD across the alignments. Applies only to RI, SVD and SGNS.

Take away Messages

- LSCD is a feasible task
- models are distributed over a wide range of performances
- OP alignment works much better than expected
- most complex model has worst performance (SCAN)
- SGNS+OP+CD is the best combination and should be the baseline for future studies
- embeddings should always be mean centered before alignment
- embeddings are more stable than expected
- be aware of frequency issues, don't use VI because of this, comparing corpora of vastly different sizes increases these issues

Open Questions

- Why does alignment (OP) work better than learning one common space (WI)?

Bibliography I

Cook, P., Lau, J. H., McCarthy, D., & Baldwin, T. (2014). Novel word-sense identification. In *25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland* (pp. 1624–1635).

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta control: Laws of semantic change and inherent biases in word representation models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1147–1156). Copenhagen, Denmark.

Frermann, L., & Lapata, M. (2016). A Bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics*, *4*, 31–45.

Lau, J. H., Cook, P., McCarthy, D., Newman, D., & Baldwin, T. (2012). Word sense induction for novel sense detection. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 591–601). Stroudsburg, PA, USA.

Schlechtweg, D., Schulte im Walde, S., & Eckmann, S. (2018). Diachronic Usage Relatedness (DURel): A Framework for the Annotation of Lexical Semantic Change. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (p. 169-174). New Orleans, Louisiana.

Takamura, H., Nagata, R., & Kawasaki, Y. (2017). Analyzing semantic change in Japanese loanwords. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 1, long papers* (pp. 1195–1204). Valencia, Spain: Association for Computational Linguistics.