University of Stuttgart
Germany

A Wind of Change: Detecting and Evaluating Lexical Semantic Change across Times and Domains

Dominik Schlechtweg Anna Hätty Marco Del Tredici Sabine Schulte im Walde

Lexical Semantic Change Detection

Diachronic LSCD: detect sense-divergences for words over time in text

(1) *1796* Ein paar **Donnerwetter** nebst Regen trugen noch mehr zur Kühle bey.
(2) *1875* Potz **Donnerwetter**, bin aber ich g'loffen!

Synchronic LSCD: from general-language to domain-specific use

(3) *general* . . . um im Winter die Gleise von **Schnee** und Eis zu befreien.
(4) *cooking* Das Eiweiss zu **Schnee** schlagen und darunterheben.

→ **we perform the first large-scale evaluation for LSC detection**

Takeaway

Representation: SGNS performs best on average

- SGNS is more stable than expected
- most complex model has low performance (SCAN) [Frermann & Lapata 2016]

Alignment: OP alignment works [Hamilton et al. 2016b]

- SGNS should be mean centered before alignment

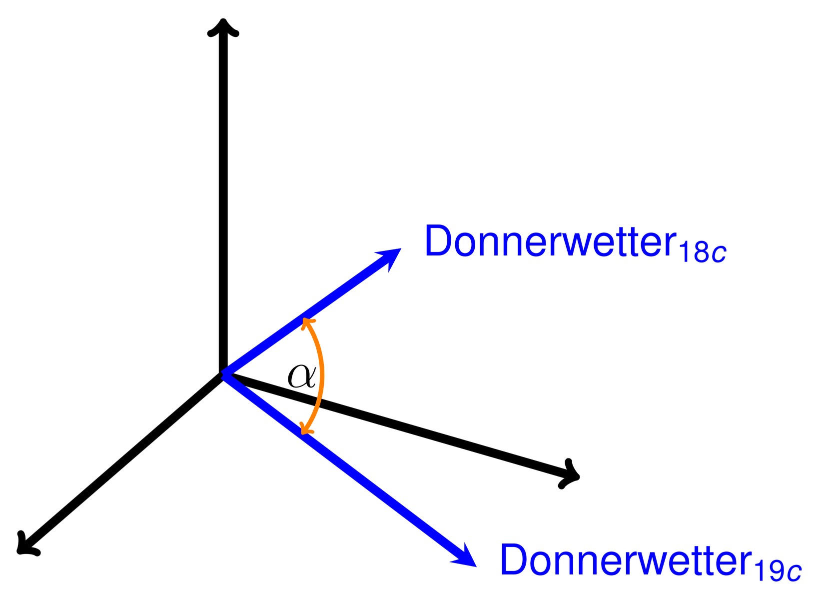
Measures: CD outperforms LND [Hamilton et al. 2016a]

- Dispersion measures have low performance [Schlechtweg et al. 2017]

Best combination: **SGNS+OP+CD**

Models

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



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

Task, Corpora & Datasets

Ranking 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.

Corpora:

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

Table: Corpora and their sizes.

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

Annotation

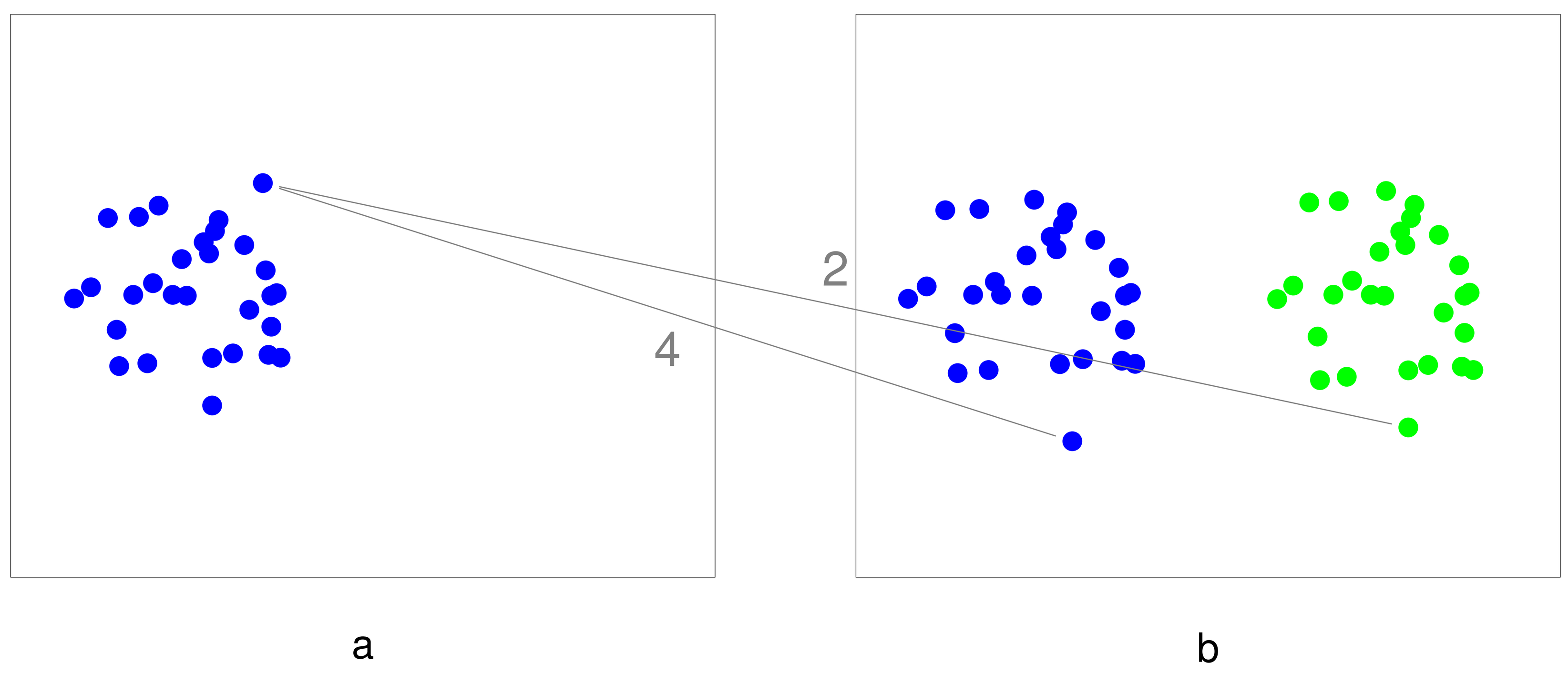


Figure: 2-dimensional use spaces in two corpora. Dots represent uses of word *w*. Spatial proximity of two uses means high relatedness.

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

Alignment:

Dataset	OP	OP_{-}	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: Mean ρ scores for CD across the alignments. Applies only to RI, SVD and SGNS.

Code for all models is available at:

`github.com/Garrafao/LSCDetection`

Acknowledgments

We thank Haim Dubossarsky, Simon Hengchen, Andres Karjus, Barbara McGillivray, Cennet Oguz, Sascha Schlechtweg, Nina Tahmasebi and the three anonymous reviewers for their valuable comments. We further thank Michael Dorna and Bingqing Wang for their helpful advice. We also thank Lea Frermann for providing the code of SCAN and helping to set up the implementation.

References

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

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). Cultural shift or linguistic drift? Comparing two computational measures of semantic change. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2116–2121). Austin, Texas.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Diachronic word embeddings reveal statistical laws of semantic change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1489–1501). Berlin, Germany.

Schlechtweg, D., Eckmann, S., Santus, E., Schulte im Walde, S., & Hole, D. (2017). German in flux: Detecting metaphoric change via word entropy. In *Proceedings of the 21st Conference on Computational Natural Language Learning* (pp. 354–367). Vancouver, Canada.