EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds

Corina Dima · Department of Linguistics and SFB 833, University of Tübingen, Germany

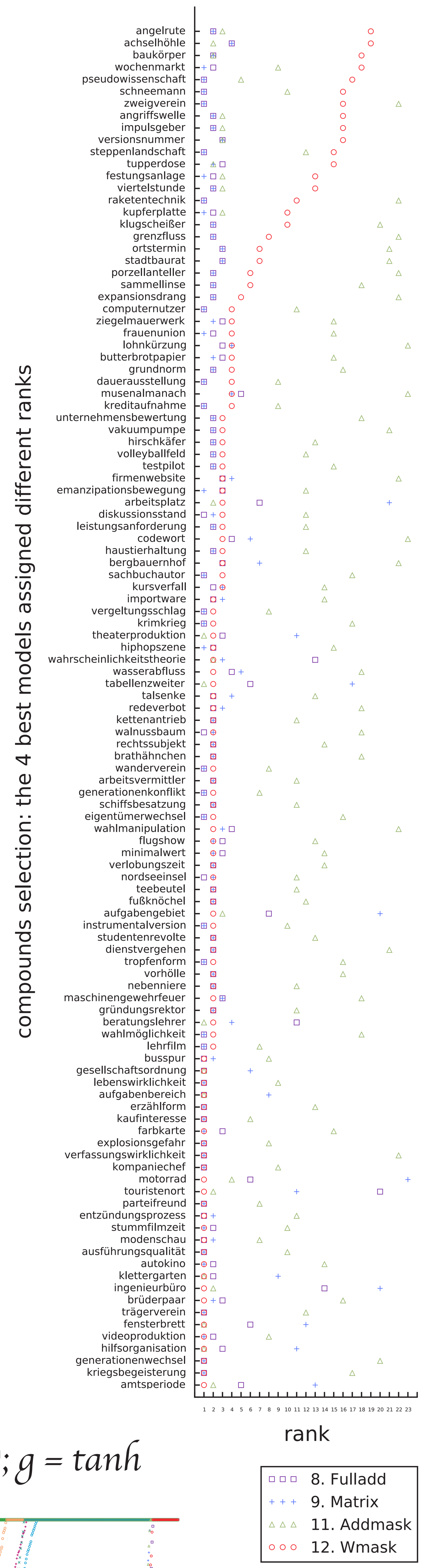
Apfel	+	Baum	→	Apfelbaum
[0.3 0.1 0.7 1.3 0.2] u		[0.5 0.9 0.1 0.4 1.2] v		[0.2 1.0 0.6 0.7 1.1] w

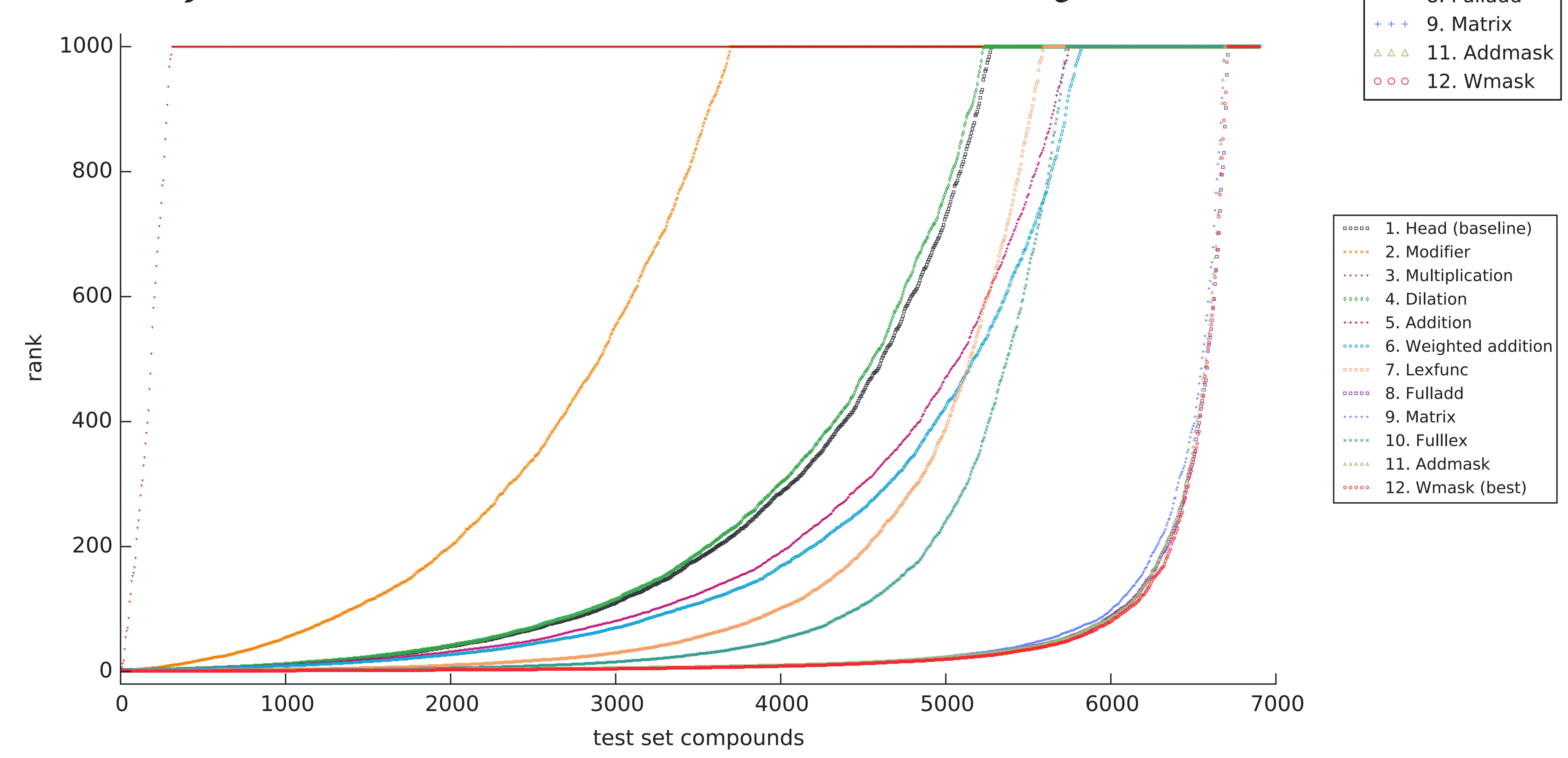
$$f\left([0.3\ 0.1\ 0.7\ 1.3\ 0.2]\,u,\ [0.5\ 0.9\ 0.1\ 0.4\ 1.2]\,v\right) = [??\ ??\ ??\ ??\ ??]\,p$$

What f makes p most similar to w?

	Model	Q1	Q2	Q3
1	$p = v$ head[1]	29	174	884
2	$p = u$ modifier[1]	144	808	≥1K
3	$p = u \odot v$ component-wise multiplication[1]	≥1K	≥1K	≥1K
4	$p = (u \cdot u)v + (\lambda-1)(u \cdot v)u$ dilation[1]	30	181	926
5	$p = 0.5u + 0.5v$ vector addition[1]	24	120	553
6	$p = \lambda u + \beta v$ weighted vector addition[1]	20	105	503
7	$p = \mathcal{U}v$ lexical function[2,3]	7	51	526.5
8	$p = M_1 u + M_2 v$ full additive[4,3]	**2**	**6**	27
9	$p = g(W[u;v])$ matrix[5]	**2**	7	29
10	$p = g(W[\mathcal{V}u; \mathcal{U}v])$ full lexical[6,3]	4	26	334
NEW! 11	$p = u \odot u' + v \odot v''$ additive mask	3	7	27
NEW! 12	$p = g(W[u \odot u'; v \odot v''])$ Wmask	**2**	**6**	**24**

where $u, v, p \in \mathbb{R}^n$; $\lambda, \beta \in \mathbb{R}$; $\mathcal{U}, \mathcal{V}, M_1, M_2 \in \mathbb{R}^{n\times n}$; $W \in \mathbb{R}^{n\times 2n}$; $g = tanh$





Task

Given a dataset of compounds together with their immediate constituents, and the corresponding distributed representations for each of the individual words, learn a *composition function* f that combines the representations of the constituents into the representation of the compound such that the *composite representation* (p) is similar to its corpus-estimated *observed representation* (w).

Results

12 composition models were evaluated on the task of building compositional representations for German compounds. The best performing model is the newly introduced *Wmask* model (model 12).

Dataset

34497 compounds from GermaNet 9.0 German compounds list; frequency filtered: modifier, head and compound with min. frequency 500 in the support corpus.

Word Representations

Trained 50, 100, 200 and 300 word representations using GloVe (Pennington et. al, 2014), a 10B token raw-text corpus extracted from the DECOW14AX corpus (Schäfer, 2015) and 1M words vocabulary.

The *mask* models

When a word w enters a composition process, there is some variation in meaning depending on whether it is the first or the second element of the composition:

company **car** (*road vehicle*)
↕
car factory (*artefact*)

The masks of a word w represented by $u \in \mathbb{R}^n$ are two vectors u' and $u'' \in \mathbb{R}^n$, initialized with a vector of all ones and estimated with the help of the training data.

Modifier masks

	0.3	0.1	0.6	0.2	1.1
u'_{car} →	0.5	0.8	0.4	0.3	0.1
	0.0	0.1	0.1	0.2	0.3
	⋮				
$v'_{factory}$ →	0.2	1.0	0.5	0.7	0.8
	0.4	0.0	0.7	0.1	0.1

Head masks

0.5	1.0	0.9	0.7	0.1	
0.2	0.0	0.3	0.6	1.4	← u''_{car}
0.9	1.3	0.6	0.3	1.1	
⋮					
0.1	0.8	0.9	0.5	0.3	← $v''_{factory}$
2.1	0.2	0.4	0.9	0.7	

Composition with the *mask* models

The composite representation of a compound like *car factory* is obtained by combining the masked representations of it's modifier and head:

$$u_{car} \odot u'_{car} \text{ and } v_{factory} \odot v''_{factory}$$

The masked representation is the result of component-wise multiplication between the initial vector of the word and the mask corresponding to its current position. The masked representations are then combined via component-wise addition (model 11) or via a global matrix $W \in \mathbb{R}^{n\times 2n}$ and a nonlinearity g (model 12).

u_{car} → [0.2 0.5 0.8 1.3 0.2]   [0.7 1.1 0.1 0.8 0.5] ← $v_{factory}$

⊙   ⊙

u'_{car} → [0.5 0.8 0.4 0.3 0.1]   [0.1 0.8 0.9 0.5 0.3] ← $v''_{factory}$

↓   ↓

$u_{car} \odot u'_{car}$ [0.1 0.4 0.3 0.4 0.0]   [0.0 0.9 0.1 0.4 0.1] $v_{factory} \odot v''_{factory}$

↓+   ↓W

[0.1 1.3 0.4 0.8 0.1]   [0.2 1.0 0.5 1.3 0.1]

$u_{car} \odot u'_{car} + v_{factory} \odot v''_{factory}$ $g(W[u_{car} \odot u'_{car}; v_{factory} \odot v''_{factory}])$

References

[1] Jeff Mitchell and Mirella Lapatta. 2010. Composition in distributional models of semantics. *Cognitive Science*, 34(8):1388-1429.
[2] Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In Proceedings of EMNLP 2010.
[3] G. Dinu, T.P. Nghia and M. Baroni. 2013. General estimation and evaluation of compositional distributional semantic models. In Proceedings of *ACL 2010 CVSC Workshop*.
[4] F. M. Zanzotto, I. Korkontzelos, F. Falluchi and S. Manandhar. 2010. Estimating Linear Models for Compositional Distributional Semantics. In *Proceedings of ICCL 2010*.
[5] R. Socher, C.D. Manning, A.Y. Ng. 2010. Learning continuous phrase representations and syntactic parsing with recursive neural networks. In *Proceedings of NIPS-2010 DL Workshop*.
[6] Richard Socher, Brody Huval, C.D. Manning and Andrew Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In Proceedings of EMNLP-CoNLL 2012.

https://github.com/corinadima/wordcomp