Collocation Classification with Unsupervised Relation Vectors

Luis Espinosa-Anke[1], Leo Wanner[2,3], and **Steven Schockaert[1]**

[1]School of Computer Science, Cardiff University, United Kingdom

[2]ICREA and [3]NLP Group, Universitat Pompeu Fabra, Barcelona, Spain

{espinosa-ankel,schockaerts1}@cardiff.ac.uk

leo.wanner@upf.edu

Abstract

Lexical relation classification is the task of predicting whether a certain relation holds between a given pair of words. In this paper, we explore to which extent the current distributional landscape based on word embeddings provides a suitable basis for classification of collocations, i.e., pairs of words between which idiosyncratic lexical relations hold. First, we introduce a novel dataset with collocations categorized according to *lexical functions*. Second, we conduct experiments on a subset of this benchmark, comparing it in particular to the well known DiffVec dataset. In these experiments, in addition to simple word vector arithmetic operations, we also investigate the role of unsupervised relation vectors as a complementary input. While these relation vectors indeed help, we also show that lexical function classification poses a greater challenge than the syntactic and semantic relations that are typically used for benchmarks in the literature.

1 Introduction

Relation classification is the task of predicting whether between a given pair of words or phrases, a certain lexical, semantic or morphosyntactic relation holds. This task has direct impact in downstream NLP tasks such as machine translation, paraphrase identification (Etzioni et al., 2005), named entity recognition (Socher et al., 2012), or knowledge base completion (Socher et al., 2013). The currently standard approach to relation classification is to combine the embeddings corresponding to the arguments of a given relation into a meaningful representation, which is then passed to a classifier. As for *which* relations have been targeted so far, the landscape is considerably more varied, although we may safely group them into morphosyntactic and semantic relations.

Morphosyntactic relations have been the focus of work on unsupervised relational similarity, as it has been shown that verb conjugation or nominalization patterns are relatively well preserved in vector spaces (Mikolov et al., 2013; Pennington et al., 2014a). Semantic relations pose a greater challenge (Vylomova et al., 2016), however. In fact, as of today, it is unclear which operation performs best (and why) for the recognition of individual lexico-semantic relations (e.g., *hyperonymy* or *meronymy*, as opposed to *cause*, *location* or *action*). Still, a number of works address this challenge. For instance, *hypernymy* has been modeled using vector concatenation (Baroni et al., 2012), vector difference and component-wise squared difference (Roller et al., 2014) as input to linear regression models (Fu et al., 2014; Espinosa-Anke et al., 2016); cf. also a sizable number of neural approaches (Shwartz et al., 2016; Anh et al., 2016). Furthermore, several high quality semantic relation datasets are available, ranging from well-known resources such as WordNet (Miller, 1995), Yago (Suchanek et al., 2007), BLESS (Baroni and Lenci, 2011), several SemEval datasets (Jurgens et al., 2012; Camacho-Collados et al., 2018) or DiffVec (Vylomova et al., 2016). But there is a surprising gap regarding collocation modeling. Collocations, which are semi-compositional in their nature in that they are situated between fixed multiword expressions (MWEs) and free (semantic) word combinations, are of relevance to second language (henceforth, L2) learners and NLP applications alike. In what follows, we investigate whether collocations can be modeled along the same lines as semantic relations between pairs of words. For this purpose, we introduce LEX-FUNC, a newly created dataset, in which collocations are annotated with respect to the semantic typology of *lexical functions* (LFs) (Mel'čuk, 1996). We use LEXFUNC to train linear SVMs on top of

different word and relation embedding composition. We show that the recognition of the semantics of a collocation, i.e., its classification with respect to the LF-typology, is a more challenging problem than the recognition of standard lexico-semantic relations, although incorporating distributional relational information brings a significant increase in performance.

2 Collocations and LᴇxFᴜɴC

We first introduce the notion of collocation and LF and then present the LᴇxFᴜɴC dataset.[1]

2.1 The phenomenon of collocation

Collocations such as *make [a] suggestion*, *attend [a] lecture*, *heavy rain*, *deep thought* or *strong tea*, to name a few, are described by Kilgarriff (2006) as *restricted lexical co-occurrences of two syntactically bound lexical items*. Due to their idiosyncrasy, collocations tend to be language-specific. For instance, in English or Norwegian we *take [a] nap*, whereas in Spanish we *throw* it, and in French, Catalan, German and Italian we *make* it. However, they are compositionally less rigid than some other types of *multiword expressions* such as, e.g., idioms (as, e.g., *[to] kick the bucket*) or multiword lexical units (as, e.g., *President of the United States* or *chief inspector*). Specifically, they are formed by a freely chosen word (the *base*), which restricts the selection of its *collocate* (e.g., *rain* restricts us to use *heavy* in English to express intensity).[2]

Recovery of collocations from corpora plays a major role in improving L2 resources, in addition to obvious advantages in NLP applications such as natural language analysis and generation, text paraphrasing / simplification, or machine translation (Hausmann, 1984; Bahns and Eldaw, 1993; Granger, 1998; Lewis and Conzett, 2000; Nesselhauf, 2005; Alonso Ramos et al., 2010).

Starting with the seminal work by Church and Hanks (1989), an extensive body of work has been produced on the detection of collocations in text corpora; cf., e.g., (Evert and Kermes, 2013; Evert, 2007; Pecina, 2008; Bouma, 2010; Garcia et al., 2017), as well as the Shared Task of the PARSEME European Cost Action on automatic recognition of verbal MWEs.[3] However, mere lists of collocations are often insufficient for both L2 acquisition and NLP. Thus, a language learner may not know the difference between, e.g., *come to fruition* and *bring to fruition* or between *have [an] approach* and *take [an] approach*, etc. Semantic labeling is required. The failure to identify the semantics of collocations also led, e.g., in earlier machine translation systems, to the necessity of the definition of collocation-specific cross-language transfer rules (Dorr, 1994; Orliac and Dillinger, 2003). The above motivates us to consider in this paper collocations and their classification in terms of LFs (Mel'čuk, 1996), their most fine-grained semantic typology (see Section 2.2). Especially because, so far, this is only discussed in a reduced number of works, and typically on a smaller scale (Wanner et al., 2006; Gelbukh and Kolesnikova., 2012).

2.2 LFs and the LᴇxFᴜɴC dataset

An LF can be viewed as a function $f(\cdot)$ that associates, with a given base L (which is the *argument* or *keyword* of f), a set of (more or less) "synonymous collocates that are selected contingent on L to manifest the meaning corresponding to f" (Mel'čuk, 1996). The name of an LF is a Latin abbreviation of this meaning. For example, Oper for *operāri* ('do', 'carry out'), Magn for *magnus* ('great', 'intense'), and so forth. The LᴇxFᴜɴC dataset consists of collocations categorized in terms of LFs. Table 1 lists the ten LFs used in this paper, along with a definition, example and frequency. The LFs have been selected so as to cover the most prominent syntactic patterns of collocations (verb+direct object, adjective+noun, and noun+noun).

3 Experimental Setup

In our experiments, we want to assess whether different LFs (i.e., semantically different collocational relations) can be captured using standard relation classification models, despite the acknowledged idiosyncratic and language-specific

LF	DEFINITION	EXAMPLE	FREQ.
magn	'very', 'intense'	*strong accent*	2,491
$oper_1$	'do', 'carry out', 'participate'	*engage* [*in an*] *activity*	1,036
$real_1$	'realize', 'accomplish', 'apply according to purpose'	*drop* [*a*] *bomb*	316
antimagn	'weak', 'little intense'	*imperceptible accent*	301
antibon	'negative', 'not as expected'	*lame attempt*	201
$causfunc_0$	'cause sth to be materialized / to function'	*lay* [*an*] *egg*	150
bon	'positive', 'as expected'	*impressive figure*	142
$liqufunc_0$	'eliminate', 'make sth. not function'	*resolve ambiguity*	118
sing	'single item or quantum of a collection or a mass'	*clove* [*of*] *garlic*	72
mult	'multitude or collection of a given item or quantum'	*bunch* [*of*] *keys*	56
TOTAL			**4,914**

Table 1: Statistics, definitions and examples of the LEXFUNC dataset. The indices indicate the argument structure of the LF: '1'stands for "first actant is the grammatical subject"; '0' for "the base is the grammatical subject".

(but still semi-compositional) "collocationality" between a collocation's *base* and *collocate*. To this end, we benchmark standard relation classification baselines in the task of LF classification. Furthermore, we also explore an explicit encoding of relational properties by distributional relation vectors (see Section 3.2).

Moreover, to contrast the LF categories in our LEXFUNC dataset with others typically found in the relation classification literature, we use ten categories from DiffVec (Vylomova et al., 2016), a dataset which was particularly designed to explore the role of vector difference in supervised relation classification. The rationale for this being that, by subtraction, the features that are common to both words are known to be "cancelled out". For instance, for **madrid − spain**, this operation can be expected to capture that the first word is a capital city and the second word is a country, and "remove" the fact that both words are related to Spain (Levy et al., 2014).

Both for DiffVec and LEXFUNC, we run experiments on those categories for which we have at least 99 instances. We cast the relation classification task as a multi-class classification problem and use a stratified $\frac{2}{3}$ portion of the data for training and the rest for evaluation. We consider each of the datasets in isolation, as well as a concatenation of both (referred to in Table 2 as DIFF-VEC+LEXFUNC). The model we use is a Linear SVM,[4] trained on a suite of vector composition

operations (Section 3.1).

3.1 Modeling relations using word vectors

Let $\mathbf{w_1}$ and $\mathbf{w_2}$ be the vector representations of two words w_1 and w_2. We experiment with the following word-level operations: DIFF ($\mathbf{w_2} - \mathbf{w_1}$), CONCAT ($\mathbf{w_1} \oplus \mathbf{w_2}$), SUM ($\mathbf{w_1} + \mathbf{w_2}$), MULT ($\mathbf{w_1} \circ \mathbf{w_2}$)), and LEFTW ($\mathbf{w_1}$), the latter operation being included to explore the degree to which the data can be lexically memorized (Levy et al., 2015)[5].

3.2 Relation vectors

Because word embeddings are limited in the amount of relational information they can capture, a number of complementary approaches have emerged which directly learn vectors that capture the relation between concepts, typically using distributional statistics from sentences mentioning both words (Espinosa-Anke and Schockaert, 2018; Washio and Kato, 2018; Joshi et al., 2018; Jameel et al., 2018). Below we explore the potential of such relation vectors for semantic relation classification. Specifically, we trained them for all word pairs from DiffVec and LEXFUNC using two different variants of the SeVeN model (Espinosa-Anke and Schockaert, 2018). The corpus for training these vectors is a Wikipedia dump from Jan-

---

[4]Implemented in *scikit-learn* (http://scikit-learn.
org.

[5]In fact, prototypicality may be a strong indicator for capturing *some* LFs. *Heavy*, for instance, may be considered as a prototypical collocate of 'magn'. However, 'heavy rain' is a more restricted English combination than 'heavy laptop', for example, but less frozen than 'heavy artillery'.

	DiffVec				DiffVec+LexFunc				LexFunc			
	P	**R**	**F**	**A**	**P**	**R**	**F**	**A**	**P**	**R**	**F**	**A**
DIFF	78.94	80.91	78.00	90.00	62.58	62.00	61.84	79.80	54.54	51.44	52.64	73.98
SUM	63.34	60.80	61.77	84.74	57.70	55.01	55.64	77.78	56.11	56.92	56.37	75.00
MULT	49.56	41.74	44.45	66.99	41.25	61.44	33.55	58.17	39.51	32.39	34.68	59.62
LEFTW	49.74	50.50	49.49	80.76	38.89	36.14	36.49	67.35	63.07	60.65	61.64	80.30
CONCAT	81.70	84.89	83.11	93.34	71.07	71.10	70.89	87.12	64.68	62.39	63.20	80.54
DIFF+RVAE	82.58	85.35	83.71	94.78	68.25	67.24	67.18	86.43	52.38	50.85	51.49	74.27
DIFF+RVAVG6	84.07	85.20	84.42	94.44	67.62	67.64	67.16	85.55	59.52	59.83	59.45	77.92
SUM+RVAE	73.14	70.60	71.49	90.25	65.56	60.67	62.82	84.67	61.54	57.31	59.13	78.21
SUM+RVAVG6	71.62	69.06	70.15	90.64	67.27	65.02	65.58	85.28	61.63	59.83	60.58	78.71
MULT+RVAE	65.27	53.57	57.59	82.95	53.86	44.42	47.15	75.28	44.43	37.22	39.36	68.22
MULT+RVAVG6	69.72	57.84	62.18	85.94	55.23	50.02	52.10	79.04	54.21	49.16	51.07	71.94
LEFTW+RVAE	65.16	61.51	62.64	90.88	50.37	46.35	47.51	80.38	45.65	39.45	41.69	66.96
LEFTW+RVAVG6	72.30	65.23	67.62	91.34	57.14	54.23	55.21	83.02	62.71	59.42	60.84	79.84
CONCAT+RVAE	86.23	87.78	86.58	95.79	72.12	72.51	72.08	89.33	64.43	60.14	61.84	80.83
CONCAT+RVAVG6	**88.09**	**88.27**	**88.09**	**95.89**	**73.65**	**74.26**	**73.73**	**89.88**	**67.23**	**68.30**	**67.70**	**81.92**

Table 2: Experimental results of several baselines on different multiclass settings for relation classification.

uary 2018, with GloVe (Pennington et al., 2014b) 300d pre-trained embeddings.

The first variant, referred to as RVAVG6, is based on averaging the words that appear in sentences mentioning the two given target words. Since this approach differentiates between words that appear before the first word, after the second word, or in between the two words, and takes into account the order in which the words appear, it results in relation vectors with a dimensionality which is six times the dimensionality of the considered word vectors. The second variant, referred to as RVAE, starts from the same high-dimensional relation vector, but then uses a conditional autoencoder to obtain a lower-dimensional and potentially higher-quality 300d vector[6].

4 Results

Table 2 shows the experimental results for Diff-Vec, LexFunc and both datasets together. The first five rows show the performance of word embedding operations, whereas the configurations for remaining rows also include a relation vector.

4.1 Discussion

We highlight two major conclusions. First, despite vector difference and component-wise multiplication being the most popular vector operations for encoding relations between words, also

---
[6]We used the code available at `bitbucket.com/luisespinosa/seven` for obtaining both representations.

in more expensive neural architectures for relation modeling (Washio and Kato, 2018; Joshi et al., 2018), vector concatenation alone proves to be a highly performing baseline. Moreover, the overall best method (CONCAT+RVAVG6), obtains performance gains when compared with the standard DIFF method ranging from +5.89% in DiffVec to +7.98% in LexFunc and +10.08% in the combined dataset. This suggests that while vector differences may encode relational properties, important information is lost when only this operation is considered.

Second, despite being a well-studied topic, recognizing lexical functions emerges as a challenging problem. They seem difficult to classify, not only between themselves, but also when coupled with other lexical semantic relations. This may be due to the fact that collocations are *idiosyncratic lexical co-occurrences* which are *syntactically bound*. The base and collocate embeddings should account for these properties, rather than over-relying on the contexts in which they appear. In the following section we present an analysis on the main sources of confusion in the LexFunc and DiffVec+LexFunc settings.

4.2 Problematic LFs

We aim to gain an understanding of recurrent errors made both by the best performing model (CONCAT+RVAVG6) and DIFF. Figure 1 shows confusion matrices for the two datasets involving LFs,
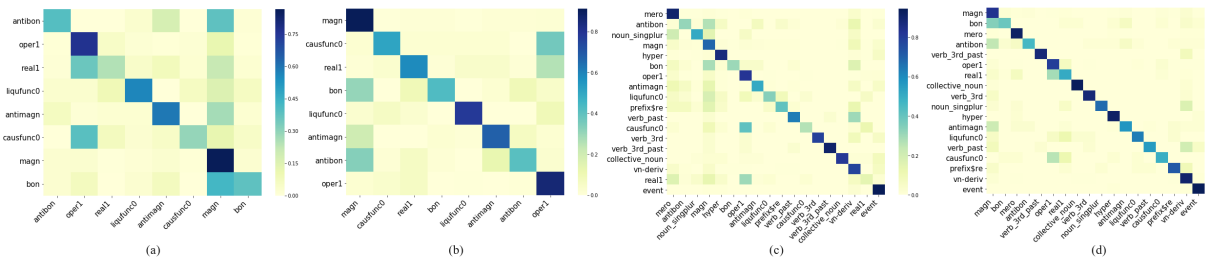
Figure 1: Confusion matrices on the best performing model (CONCAT+RVAVG6): (1a) and (1b) show performance on the LEXFUNC experiment, and (1c) and (1d) on DIFFVEC+LEXFUNC.

namely LEXFUNC and DIFFVEC+LEXFUNC. We are particularly interested in pinpointing which LFs are most difficult to classify, and whether there is any particular label that agglutinates most predictions. For example, in (Fig. 1a) we see a strong source of confusion in the DIFF model between the 'bon' and 'magn' labels. Both are noun-adjective combinations and both are used as intensifiers, but they subtly differ in that only one enforces a perceived degree of positiveness (e.g., *resounding* vs. *crushing victory*). Thus, combining their vectors produces clearly similar representations that confuse the classifier, a scenario which is only partly alleviated by the use of relation vectors (Fig. 1b).

The case of 'oper$_1$' (perform) and 'real$_1$' (accomplish) also proves problematic. The number of light verbs as collocates of these LFs is notably high in the former, amounting to 48%; 'real$_1$' is more semantic, with almost 11% light verbs. Interestingly, however, these labels are almost never confused with the 'event' label from DiffVec (Figs. 1c and 1d), even if it also contains relations with light verbs such as *break* or *pay*.

Finally, one last source of confusion that warrants discussion involves 'magn' and 'antimagn', two noun-advective collocations which are different in that the former conveys a notion of intensity, whereas the latter is about weakness (e.g., 'faint admiration' or 'slight advantage'). These two LFs typically include antonymic collocates (e.g., 'weak' and 'strong' as collocates for the base 'argument'), and these are known to have similar distributional vectors (Mrkšić et al., 2016; Nguyen et al., 2016), which in high likelihood constitutes a source of confusion.

5 Conclusions and Future Work

In this paper, we have discussed the task of distributional collocation classification. We have used a set of collocations categorized by *lexical functions*, as introduced in the Meaning Text Theory (Mel'čuk, 1996), and evaluated a wide range of vector representations of relations. In addition, we have used the DiffVec (Vylomova et al., 2016) dataset to provide a frame of reference, as this dataset has been extensively studied in the distributional semantics literature, mostly for evaluating the role of vector difference. We found that, despite this operation being the go-to representation for lexical relation modeling, concatenation works as well or better, and clear improvements can be obtained by incorporating explicitly learned relation vectors. However, even with these improvements, categorizing LFs proves to be a difficult task.

In the future, we would like to experiment with more data, so that enough training data can be obtained for less frequent LFs. To this end, we could benefit from the supervised approach proposed in (Rodríguez-Fernández et al., 2016), and then filter by pairwise correlation strength metrics such as PMI. Another exciting avenue would involve exploring cross-lingual transfer of LFs, taking advantage of recent development in unsupervised cross-lingual embedding learning (Artetxe et al., 2017; Conneau et al., 2017).

References

M. Alonso Ramos, L. Wanner, O. Vincze, G. Casamayor, N. Vázquez, E. Mosqueira, and S. Prieto. 2010. Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, pages 3209–3214, La Valetta, Malta.

Tuan Luu Anh, Yi Tay, Siu Cheung Hui, and See Kiong Ng. 2016. Learning term embeddings for taxonomic relation identification using dynamic weighting neural network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 403–413.

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2017. Unsupervised neural machine translation. *arXiv preprint arXiv:1710.11041*.

J. Bahns and M. Eldaw. 1993. Should we teach EFL students collocations? *System*, 21(1):101–114.

Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment above the word level in distributional semantics. In *Proceedings of EACL*, pages 23–32.

Marco Baroni and Alessandro Lenci. 2011. How we blessed distributional semantic evaluation. In *Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 1–10. Association for Computational Linguistics.

M. Benson. 1989. The structure of the collocational dictionary. *International Journal of Lexicography*, 2(1):1–13.

J. Binon and S. Verlinde. 2013. Electronic pedagogical dictionaries. In R. Gouws, U. Heid, W. Schweickard, and H.E. Wiegand, editors, *Dictionaries. An International Encyclopedia of Lexicography*, pages 1035–1046. De Gruyter Mouton, Berlin.

G. Bouma. 2010. Collocation extraction beyond the independence assumption. In *Proceedings of the ACL 2010, Short paper track*, Uppsala.

Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. 2018. Semeval-2018 task 9: hypernym discovery. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 712–724.

K. Church and P. Hanks. 1989. Word Association Norms, Mutual Information, and Lexicography. In *Proceedings of the 27th Annual Meeting of the ACL*, pages 76–83.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

A. Cowie. 1994. Phraseology. In R.E. Asher and J.M.Y. Simpson, editors, *The Encyclopedia of Language and Linguistics, Vol. 6*, pages 3168–3171. Pergamon, Oxford.

B.J. Dorr. 1994. Machine translation divergences: A formal description and proposed solution. *Computational linguistics*, 20:579–634.

Luis Espinosa-Anke, Jose Camacho-Collados, Claudio Delli Bovi, and Horacio Saggion. 2016. Supervised distributional hypernym discovery via domain adaptation. In *Conference on Empirical Methods in Natural Language Processing; 2016 Nov 1-5; Austin, TX. Red Hook (NY): ACL; 2016. p. 424-35*. ACL (Association for Computational Linguistics).

Luis Espinosa-Anke and Steven Schockaert. 2018. Seven: Augmenting word embeddings with unsupervised relation vectors. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2653–2665.

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S Weld, and Alexander Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. *Artificial intelligence*, 165(1):91–134.

S. Evert. 2007. Corpora and collocations. In A. Lüdeling and M. Kytö, editors, *Corpus Linguistics. An International Handbook*. Mouton de Gruyter, Berlin.

S. Evert and H. Kermes. 2013. Experiments on candidate data for collo- cation extraction. In *Companion Volume to the Proceedings of the 10th Conference of the EACL*, pages 472–487.

Ruiji Fu, Jiang Guo, Bing Qin, Wanxiang Che, Haifeng Wang, and Ting Liu. 2014. Learning semantic hierarchies via word embeddings. In *Proceedings of ACL*, volume 1.

Marcos Garcia, Marcos García-Salido, and Margarita Alonso-Ramos. 2017. Using bilingual word-embeddings for multilingual collocation extraction. In *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, pages 21–30.

A. Gelbukh and O. Kolesnikova. 2012. *Semantic Analysis of Verbal Collocations with Lexical Functions*. Springer, Heidelberg.

S. Granger. 1998. Prefabricated patterns in advanced EFL writing: Collocations and Formulae. In A. Cowie, editor, *Phraseology: Theory, Analysis and Applications*, pages 145–160. Oxford University Press, Oxford.

F.J. Hausmann. 1984. Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortwendungen. *Praxis des neusprachlichen Unterrichts*, 31(1):395–406.

Shoaib Jameel, Zied Bouraoui, and Steven Schockaert. 2018. Unsupervised learning of distributional relation vectors. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 23–33.

Mandar Joshi, Eunsol Choi, Omer Levy, Daniel S Weld, and Luke Zettlemoyer. 2018. pair2vec: Compositional word-pair embeddings for cross-sentence inference. *arXiv preprint arXiv:1810.08854*.

David A Jurgens, Peter D Turney, Saif M Mohammad, and Keith J Holyoak. 2012. Semeval-2012 task 2: Measuring degrees of relational similarity. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 356–364. Association for Computational Linguistics.

A. Kilgarriff. 2006. Collocationality (and how to measure it). In *Proceedings of the Euralex Conference*, pages 997–1004, Turin, Italy. Springer-Verlag.

Omer Levy, Yoav Goldberg, and Israel Ramat-Gan. 2014. Linguistic regularities in sparse and explicit word representations. In *Proceedings of CoNLL*, pages 171–180.

Omer Levy, Steffen Remus, Chris Biemann, Ido Dagan, and Israel Ramat-Gan. 2015. Do supervised distributional methods really learn lexical inference relations? In *Proceedings of NAACL 2015*, Denver, Colorado, USA.

M. Lewis and J. Conzett. 2000. *Teaching Collocation. Further Developments in the Lexical Approach.* LTP, London.

I. Mel'čuk. 1995. Phrasemes in Language and Phraseology in Linguistics. In M. Everaert, E.-J. van der Linden, A. Schenk, and R. Schreuder, editors, *Idioms: Structural and Psychological Perspectives*, pages 167–232. Lawrence Erlbaum Associates, Hillsdale.

I.A. Mel'čuk. 1996. Lexical functions: A tool for the description of lexical relations in the lexicon. In L. Wanner, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–102. Benjamins Academic Publishers, Amsterdam.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

George A Miller. 1995. WordNet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Nikola Mrkšić, Diarmuid O Séaghdha, Blaise Thomson, Milica Gašić, Lina Rojas-Barahona, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young. 2016. Counter-fitting word vectors to linguistic constraints. *arXiv preprint arXiv:1603.00892*.

N. Nesselhauf. 2005. *Collocations in a Learner Corpus*. Benjamins Academic Publishers, Amsterdam.

Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. 2016. Integrating distributional lexical contrast into word embeddings for antonym-synonym distinction. *arXiv preprint arXiv:1605.07766*.

Brigitte Orliac and Mike Dillinger. 2003. Collocation extraction for machine translation. In *Proceedings of Machine Translation Summit IX*, pages 292–298.

P. Pecina. 2008. A machine learning approach to multiword expression extraction. In *Proceedings of the LREC 2008 Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 54–57, Marrakech.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014a. GloVe: Global vectors for word representation. In *Proceedings of EMNLP*, pages 1532–1543.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014b. Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*, 12.

Sara Rodríguez-Fernández, Luis Espinosa Anke, Roberto Carlini, and Leo Wanner. 2016. Semantics-driven recognition of collocations using word embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 499–505.

Stephen Roller, Katrin Erk, and Gemma Boleda. 2014. Inclusive yet selective: Supervised distributional hypernymy detection. In *Proceedings of COLING*, pages 1025–1036.

Vered Shwartz, Yoav Goldberg, and Ido Dagan. 2016. Improving hypernymy detection with an integrated path-based and distributional method. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2389–2398.

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 joint conference on empirical*

methods in natural language processing and computational natural language learning, pages 1201–1211. Association for Computational Linguistics.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: A core of semantic knowledge. unifying WordNet and Wikipedia. pages 697–706.

Ekaterina Vylomova, Laura Rimell, Trevor Cohn, and Timothy Baldwin. 2016. Take and took, gaggle and goose, book and read: Evaluating the utility of vector differences for lexical relation learning. In *Proceedings of ACL*.

L. Wanner, B. Bohnet, and M. Giereth. 2006. Making sense of collocations. *Computer Speech and Language*, 20(4):609–624.

Koki Washio and Tsuneaki Kato. 2018. Neural latent relational analysis to capture lexical semantic relations in a vector space. *arXiv preprint arXiv:1809.03401*.