Cross-Lingual Syntactically Informed Distributed Word Representations

Ivan Vulić

Language Technology Lab
DTAL, University of Cambridge
`iv250@cam.ac.uk`

Abstract

We develop a novel cross-lingual word representation model which injects syntactic information through dependency-based contexts into a shared cross-lingual word vector space. The model, termed CL-DEPEMB, is based on the following assumptions: (1) dependency relations are largely language-independent, at least for related languages and prominent dependency links such as direct objects, as evidenced by the Universal Dependencies project; (2) word translation equivalents take similar grammatical roles in a sentence and are therefore substitutable within their syntactic contexts. Experiments with several language pairs on word similarity and bilingual lexicon induction, two fundamental semantic tasks emphasising semantic similarity, suggest the usefulness of the proposed syntactically informed cross-lingual word vector spaces. Improvements are observed in both tasks over standard cross-lingual "offline mapping" baselines trained using the same setup and an equal level of bilingual supervision.

1 Introduction

In recent past, NLP as a field has seen tremendous utility of *distributed word representations* (or word embeddings, termed WEs henceforth) as features in a variety of downstream tasks (Turian et al., 2010; Collobert et al., 2011; Baroni et al., 2014; Chen and Manning, 2014). The quality of these representations may be further improved by leveraging cross-lingual (CL) distributional information, as evidenced by the recent body of work focused on learning *cross-lingual word embeddings* (Klementiev et al., 2012; Zou et al., 2013; Hermann and

Blunsom, 2014; Gouws et al., 2015; Coulmance et al., 2015; Duong et al., 2016, inter alia).[1] The inclusion of cross-lingual information results in a *shared cross-lingual word vector space* (SCLVS), which leads to improvements on monolingual tasks (typically word similarity) (Faruqui and Dyer, 2014; Rastogi et al., 2015; Upadhyay et al., 2016), and also supports cross-lingual tasks such as bilingual lexicon induction (Mikolov et al., 2013a; Gouws et al., 2015; Duong et al., 2016), cross-lingual information retrieval (Vulić and Moens, 2015; Mitra et al., 2016), entity linking (Tsai and Roth, 2016), and cross-lingual knowledge transfer for resource-lean languages (Søgaard et al., 2015; Guo et al., 2016).

Another line of work has demonstrated that syntactically informed dependency-based (DEPS) word vector spaces in monolingual settings (Lin, 1998; Padó and Lapata, 2007; Utt and Padó, 2014) are able to capture finer-grained distinctions compared to vector spaces based on standard bag-of-words (BOW) contexts. Dependency-based vector spaces steer the induced WEs towards functional similarity (e.g., *tiger:cat*) rather than topical similarity/relatedness (e.g., *tiger:jungle*), They support a variety of similarity tasks in monolingual settings, typically outperforming BOW contexts for English (Bansal et al., 2014; Hill et al., 2015; Melamud et al., 2016). However, despite the steadily growing landscape of CL WE models, each requiring a different form of cross-lingual supervision to induce a SCLVS, *syntactic information* is still typically discarded in the SCLVS learning process.

To bridge this gap, in this work we develop a new cross-lingual WE model, termed CL-DEPEMB, which *injects syntactic information into a SCLVS*. The model is supported by the recent initiatives on language-agnostic annotations for *universal lan-*

---

[1] For a comprehensive overview of cross-lingual word embedding models, we refer the reader to two recent survey papers (Upadhyay et al., 2016; Vulić and Korhonen, 2016b).

guage processing (i.e., universal POS (UPOS) tagging and dependency (UD) parsing) (Nivre et al., 2015). Relying on cross-linguistically consistent UD-typed dependency links in two languages plus a word translation dictionary, the model assumes that one-to-one word translations are substitutable within their syntactic contexts in both languages. It constructs hybrid cross-lingual dependency trees which could be used to extract monolingual and cross-lingual dependency-based contexts (further discussed in Sect. 2 and illustrated by Fig. 1).

In summary, our focused contribution is a new syntactically informed cross-lingual WE model which takes advantage of the normalisation provided by the Universal Dependencies project to facilitate the syntactic mapping across languages. We report results on two semantic tasks, monolingual word similarity (WS) and bilingual lexicon induction (BLI), which evaluate the monolingual and cross-lingual quality of the induced SCLVS. We observe consistent improvements over baseline CL WE models which require the same level of bilingual supervision (i.e., a word translation dictionary). For this supervision setting, we show a clear benefit of joint *online* training compared to standard *offline* models which construct two separate monolingual BOW-based or DEPS-based WE spaces, and then map them into a SCLVS using dictionary entries as done in (Mikolov et al., 2013a; Dinu et al., 2015; Lazaridou et al., 2015; Vulić and Korhonen, 2016b, inter alia)

2 Methodology

Representation Model   In all experiments, we opt for a standard and robust choice in vector space modeling: skip-gram with negative sampling (SGNS) (Mikolov et al., 2013b; Levy et al., 2015). We use `word2vecf`, a reimplementation of `word2vec` which is capable of learning from arbitrary $(word, context)$ pairs[2], thus clearly emphasising the role of context in WE learning.

(Universal) Dependency-Based Contexts   A standard procedure to extract dependency-based contexts (DEPS) (Padó and Lapata, 2007; Utt and Padó, 2014) from monolingual data is as follows. Given a parsed training corpus, for each target w with modifiers $m_1, \ldots, m_k$ and a head h, w is paired with context elements

$m_1\_r_1, \ldots, m_k\_r_k, h\_r_h^{-1}$, where r is the type of the dependency relation between the head and the modifier (e.g., `amod`), and r^{-1} denotes an inverse relation.[3] When extracting DEPS, we adopt the post-parsing prepositional arc collapsing procedure (Levy and Goldberg, 2014a) (see Fig. 1a-1b).

Cross-Lingual DEPS: CL-DEPEMB   First, a UD-parsed monolingual training corpus is obtained in both languages L_1 and L_2. The use of the interlingual UD scheme enables linking dependency trees in both languages (see the structural similarity of the two sentences in English (EN) and Italian (IT), Fig. 1a-1b). For instance, the link between EN words *Australian* and *scientist* as well as IT words *australiano* and *scienzato* is typed `amod` in both trees. This link generates the following monolingual EN DEPS: *(scientist, Australian_amod), (Australian, scientist_amod^{-1})* (similar for IT).

Now, assume that we possess an EN-IT translation dictionary D with pairs $[w_1, w_2]$ which contains entries *[Australian, australiano]* and *[scientist, scienzato]*. Given the observed similarity in the sentence structure, and the fact that words from a translation pair tend to take similar UPOS tags and similar grammatical roles in a sentence, we can substitute w_1 with w_2 in all DEPS in which w_1 participates (and vice versa, replace w_2 with w_1). Using the substitution idea, besides the original monolingual EN and IT DEPS contexts, we now generate additional hybrid cross-lingual EN-IT DEPS contexts: *(scientist, australiano_amod), (australiano, scientist_amod^{-1}), (scienzato, Australian_amod), (Australian, scienzato_amod^{-1})* (again, we can also generate such hybrid IT-EN DEPS contexts).

CL-DEPEMB then trains *jointly* on such extended DEPS contexts containing both monolingual and cross-lingual *(word, context)* dependency-based pairs. With CL-DEPEMB, words are considered similar if they often co-occur with similar words (and their translations) in the same dependency relations in both languages. For instance, words *discovers* and *scopre* might be considered similar as they frequently co-occur as predicates for the nominal subjects (`nsubj`) *scientist* and *scienzato*, and *stars* and *stelle* are their frequent direct objects (`dobj`). An illustrative example of the core idea behind CL-DEPEMB is provided in Fig. 1.

[3]Given an example from Fig. 1, the DEPS contexts of *discovers* are: *scientist_nsubj, stars_dobj, telescope_nmod*. Compared to BOW, DEPS capture longer-range relations (e.g., *telescope*) and filter out "accidental contexts" (e.g., *Australian*).
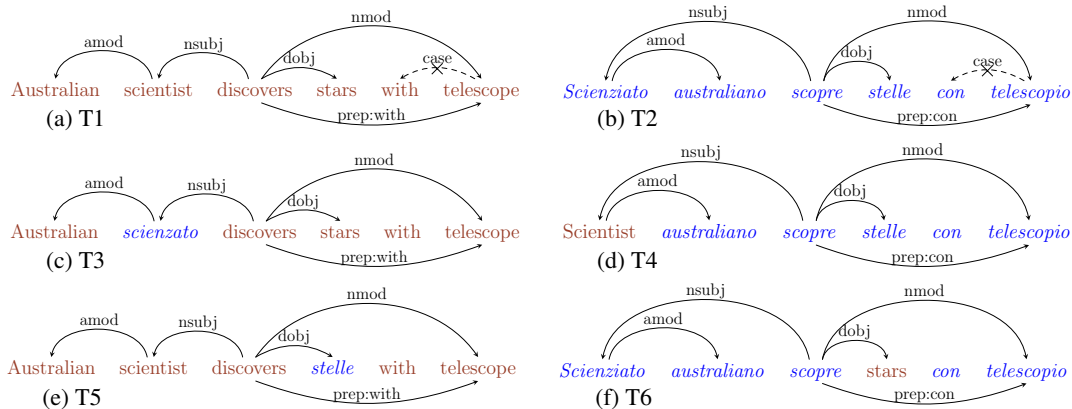
Figure 1: An example of extracting mono and CL DEPS contexts from UD parses in EN and IT assuming two dictionary entries *[scientist, scienzato]*, *[stars, stelle]*. **(T1)**: the example EN sentence taken from (Levy and Goldberg, 2014a), UD-parsed. **(T2)**: the same sentence in IT, UD-parsed; Note the very similar structure of the two parses and the use of prepositional arc collapsing (e.g., the typed link *prep_with*). **(T3)**: the hybrid EN-IT dependency tree where the EN word *scientist* is replaced by its IT translation *scienzato*. **(T4)**: the hybrid IT-EN tree using the same translation pair. **(T5)** and **(T6)**: the hybrid EN-IT and IT-EN trees obtained using the lexicon entry *(stars, stelle)*. While monolingual dependency-based representation models use only monolingual trees T1 and T2 for training, our CL-DEPEMB model additionally trains on the (parts of) hybrid trees T3-T6, combining monolingual $(word, context)$ training examples with cross-lingual training examples such as *(discovers, stelle_dobj)* or *(australiano, scientist_amod^{-1})*. Although the two sentences (T1 and T2) are direct translations of each other for illustration purposes, we stress that the proposed CL-DEPEMB model does not assume the existence of parallel data nor requires it.

Offline Models vs CL-DEPEMB (Joint) CL-DEPEMB uses a dictionary D as the bilingual signal to tie two languages into a SCLVS. A standard CL WE learning scenario in this setup is as follows (Mikolov et al., 2013a; Vulić and Korhonen, 2016b): (1) two separate monolingual WE spaces are induced using SGNS; (2) dictionary entries from D are used to learn a mapping function mf from the L_1 space to the L_2 space; (3) when mf is applied to all L_1 word vectors, the transformed L_1 space together with the L_2 space is a SCLVS. Monolingual WE spaces may be induced using different context types (e.g., BOW or DEPS). Since the transformation is done after training, these models are typically termed *offline* CL WE models.

On the other hand, given a dictionary link $[w_1, w_2]$, between an L_1 word w_1 and an L_2 word w_2, our CL-DEPEMB model performs an *online* training: it uses the word w_1 to predict syntactic neighbours of the word w_2 and vice versa. In fact, we train a single SGNS model with a joint vocabulary on two monolingual UD-parsed datasets with additional cross-lingual dependency-based training examples fused with standard monolingual DEPS pairs. From another perspective, the CL-DEPEMB model trains an extended dependency-based SGNS

model now composed of four joint SGNS models between the following language pairs: $L_1 \rightarrow L_1$, $L_1 \rightarrow L_2$, $L_2 \rightarrow L_1$, $L_2 \rightarrow L_2$ (see Fig. 1).[4]

3 Experimental Setup

We report results with two language pairs: English-German/Italian (EN-DE/IT) due to the availability of comprehensive test data for these pairs (Leviant and Reichart, 2015; Vulić and Korhonen, 2016a).

Training Setup and Parameters For all languages, we use the Polyglot Wikipedia data (Al-Rfou et al., 2013).[5] as monolingual training data. All corpora were UPOS-tagged and UD-parsed using the procedure of Vulić and Korhonen (2016a): UD treebanks v1.4, TurboTagger for tagging (Martins et al., 2013), Mate Parser v3.61 with suggested settings (Bohnet, 2010).[6] The SGNS preprocessing scheme is standard (Levy and Goldberg, 2014a):

---

[4]A similar idea of *extended joint CL training* was discussed previously by (Luong et al., 2015; Coulmance et al., 2015). In this work, we show that expensive parallel data and word alignment links are not required to produce a SCLVS. Further, instead of using BOW contexts, we demonstrate how to use DEPS contexts for joint training in the CL settings.

[5]https://sites.google.com/site/rmyeid/projects/polyglot

[6]LAS scores on the TEST portion of each UD treebank are: 0.852 (EN), 0.884 (IT), 0.802 (DE).

all tokens were lowercased, and words and contexts that appeared less than 100 times were filtered out.[7] We report results with $d = 300$-dimensional WEs, as similar trends are observed with other d-s.

Implementation The code for generating monolingual and cross-lingual dependency-based *(word, context)* pairs for the `word2vecf` SGNS training using a bilingual dictionary D is available at: `https://github.com/cambridgeltl/cl-depemb/`.

Translation Dictionaries We report results with a dictionary D labelled BNC+GT: a list of 6,318 most frequent EN lemmas in the BNC corpus (Kilgarriff, 1997) translated to DE and IT using Google Translate (GT), and subsequently cleaned by native speakers. A similar setup was used by (Mikolov et al., 2013a; Vulić and Korhonen, 2016b). We also experiment with `dict.cc`, a freely available large online dictionary (`http://www.dict.cc/`), and find that the relative model rankings stay the same in both evaluation tasks irrespective to the chosen D.

Baseline Models CL-DEPEMB is compared against two relevant *offline* models which also learn using a seed dictionary D: (1) OFF-BOW2 is a linear mapping model from (Mikolov et al., 2013a; Dinu et al., 2015; Vulić and Korhonen, 2016b) which trains two SGNS models with the window size 2, a standard value (Levy and Goldberg, 2014a); we also experiment with more informed positional BOW contexts (Schütze, 1993; Levy and Goldberg, 2014b) (OFF-POSIT2); (2) OFF-DEPS trains two DEPS-based monolingual WE spaces and linearly maps them into a SCLVS. Note that OFF-DEPS uses exactly the same information (i.e., UD-parsed corpora plus dictionary D) as CL-DEPEMB.

4 Results and Discussion

Evaluation Tasks Following Luong et al. (2015) and Duong et al. (2016), we argue that good cross-lingual word representations should preserve both monolingual and cross-lingual representation quality. Therefore, similar to (Duong et al., 2016; Upadhyay et al., 2016), we test cross-lingual WEs in two core semantic tasks: *monolingual word similarity* (WS) and *bilingual lexicon induction* (BLI).

Model	IT		DE		EN (with IT)	
	All — Verbs		All — Verbs		All — Verbs	
MONO-SGNS	0.235 — 0.318		0.305 — 0.259		0.331 — 0.281	
OFF-BOW2	0.254 — 0.317		**0.306** — 0.263		0.328 — 0.279	
OFF-POSIT2	0.227 — 0.323		0.283 — 0.194		0.336 — **0.316**	
OFF-DEPS	0.199 — 0.308		0.258 — 0.214		0.334 — 0.311	
CL-DEPEMB	**0.287 — 0.358**		**0.306 — 0.319**		**0.356** — 0.308	

Table 1: WS results on multilingual SimLex-999. All scores are Spearman's ρ correlations. MONO-SGNS refers to the best scoring monolingual SGNS model in each language (BOW2, POSIT2 or DEPS). *Verbs* refers to the verb subset of each SimLex-999.

Model	IT-EN		DE-EN	
	SL-TRANS	VULIC1K	SL-TRANS	UP1328
OFF-BOW2	0.328 [0.457]	0.405	0.218 [0.246]	0.317
OFF-POSIT2	0.219 [0.242]	0.272	0.115 [0.056]	0.185
OFF-DEPS	0.169 [0.065]	0.271	0.108 [0.051]	0.162
CL-DEPEMB	**0.541 [0.597]**	**0.532**	**0.503 [0.385]**	**0.436**

Table 2: BLI results (*Top 1* scores). For SL-TRANS we also report results on the verb translation subtask (numbers in square brackets).

Word Similarity Word similarity experiments were conducted on the benchmarking multilingual SimLex-999 evaluation set (Leviant and Reichart, 2015) which provides monolingual similarity scores for 999 word pairs in English, German, and Italian.[8] The results for the three languages are displayed in Tab. 1.

These results suggest that CL-DEPEMB is the best performing and most robust model in our comparison across all three languages, providing the first insight that the online training with the extended set of DEPS pairs is indeed beneficial for modeling true (functional) similarity.

We also carry out tests in English using another word similarity metric: QVEC,[9] which measures how well the induced word vectors correlate with a matrix of features from manually crafted lexical resources and is better aligned with downstream performance (Tsvetkov et al., 2015). The results are again in favour of CL-DEPEMB with a QVEC score of 0.540 (BNC+GT) and 0.543 (`dict.cc`), compared to those of OFF-BOW2 (0.496), OFF-POSIT2 (0.510), and OFF-DEPS (0.528).

Bilingual Lexicon Induction BLI experiments were conducted on several standard test sets: IT-

---

[7]Exactly the same vocabularies were used with all models ($\sim$ 185K distinct EN words, 163K DE words, and 83K IT words). All `word2vecf` SGNS models were trained using standard settings: 15 epochs, 15 negative samples, global

(decreasing) learning rate 0.025, subsampling rate $1e - 4$.

[8]http://technion.ac.il/~ira.leviant/MultilingualVSMdata.html
[9]https://github.com/ytsvetko/qvec

OFF-DEPS	0.259
BEST-BASELINE	0.271
CL-DEPEMB (+IT)	0.285
CL-DEPEMB (+DE)	**0.310**

Table 3: WS EN results on SimVerb-3500 (Spearman's ρ correlation scores). BEST-BASELINE refers to the best score across all baseline modeling variants. We report results of CL-DEPEMB with `dict.cc` after multilingual training with Italian (+IT) and German (+DE).

EN was evaluated on VULIC1K (Vulić and Moens, 2013a), containing 1,000 IT nouns and their EN translations, and DE-EN was evaluated on UP1328 (Upadhyay et al., 2016), containing 1,328 test pairs of mixed POS tags. In addition, we evaluate both language pairs on SimLex-999 word translations (Leviant and Reichart, 2015), containing $\sim$ 1K test pairs (SL-TRANS). We report results using a standard BLI metric: *Top 1* scores. The same trends are visible with *Top 5* and *Top 10* scores. All test word pairs were removed from D for training.

The results are summarised in Tab. 2, indicating significant improvements with CL-DEPEMB (McNemar's test, $p < 0.05$). The gap between the online CL-DEPEMB model and the offline baselines is now even more prominent,[10] and there is a huge difference in performance between OFF-DEPS and CL-DEPEMB, two models using exactly the same information for training.

Experiments on Verbs Following prior work, e.g., (Bansal et al., 2014; Melamud et al., 2016; Schwartz et al., 2016), we further show that WE models which capture functional similarity are especially important for modelling particular "more grammatical" word classes such as verbs and adjectives. Therefore, in Tab. 1 and Tab. 2 we also report results on verb similarity and translation. The results indicate that injecting syntax into cross-lingual word vector spaces leads to clear improvements on modelling verbs in both evaluation tasks.

We further verify the intuition by running experiments on another word similarity evaluation set, which targets verb similarity in specific: SimVerb-3500 (Gerz et al., 2016) contains similarity scores for 3,500 verb pairs. The results of the CL-

---

[10]We also experimented with other language pairs represented in VULIC1K (Spanish/Dutch-English) and UP1328 (French/Swedish-English). The results also show similar improvements with CL-DEPEMB, not reported for brevity.

DEPEMB on SimVerb-3500 with `dict.cc` are provided in Tab. 3, further indicating the usefulness of syntactic information in multilingual settings for improved verb representations.

Similar trends are observed with *adjectives*: e.g., CL-DEPEMB with `dict.cc` obtains a ρ correlation score of 0.585 on the adjective subset of DE SimLex while the best baseline score is 0.417; for IT these scores are 0.334 vs. 0.266.

5 Conclusion and Future Work

We have presented a new cross-lingual word embedding model which injects syntactic information into a cross-lingual word vector space, resulting in improved modeling of functional similarity, as evidenced by improvements on word similarity and bilingual lexicon induction tasks for several language pairs. More sophisticated approaches involving the use of more accurate dependency parsers applicable across different languages (Ammar et al., 2016), selection and filtering of reliable dictionary entries (Peirsman and Padó, 2010; Vulić and Moens, 2013b; Vulić and Korhonen, 2016b), and more sophisticated approaches to constructing hybrid cross-lingual dependency trees (Fig. 1) may lead to further advances in future work. Other cross-lingual semantic tasks such as lexical entailment (Mehdad et al., 2011; Vyas and Carpuat, 2016) or lexical substitution (Mihalcea et al., 2010) may also benefit from syntactically informed cross-lingual representations. We also plan to test the portability of the proposed framework, relying on the abstractive assumption of language-universal dependency structures, to more language pairs, including the ones outside the Indo-European language family.

References

Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2013. Polyglot: Distributed word representations for multilingual NLP. In *CoNLL*, pages 183–192.

Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah Smith. 2016. Many languages, one parser. *Transactions of the ACL*, 4:431–444.

Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2014. Tailoring continuous word representations for dependency parsing. In *ACL*, pages 809–815.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *ACL*, pages 238–247.

Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *COLING*, pages 89–97.

Danqi Chen and Christopher D. Manning. 2014. A fast and accurate dependency parser using neural networks. In *EMNLP*, pages 740–750.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12:2493–2537.

Jocelyn Coulmance, Jean-Marc Marty, Guillaume Wenzek, and Amine Benhalloum. 2015. Trans-gram, fast cross-lingual word embeddings. In *EMNLP*, pages 1109–1113.

Georgiana Dinu, Angeliki Lazaridou, and Marco Baroni. 2015. Improving zero-shot learning by mitigating the hubness problem. In *ICLR Workshop Papers*.

Long Duong, Hiroshi Kanayama, Tengfei Ma, Steven Bird, and Trevor Cohn. 2016. Learning crosslingual word embeddings without bilingual corpora. In *EMNLP*, pages 1285–1295.

Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. In *EACL*, pages 462–471.

Daniela Gerz, Ivan Vulić, Felix Hill, Roi Reichart, and Anna Korhonen. 2016. SimVerb-3500: A large-scale evaluation set of verb similarity. In *EMNLP*, pages 2173–2182.

Stephan Gouws, Yoshua Bengio, and Greg Corrado. 2015. BilBOWA: Fast bilingual distributed representations without word alignments. In *ICML*, pages 748–756.

Jiang Guo, Wanxiang Che, David Yarowsky, Haifeng Wang, and Ting Liu. 2016. A distributed representation-based framework for cross-lingual transfer parsing. *Journal of Artificial Intelligence Research*, 55:995–1023.

Karl Moritz Hermann and Phil Blunsom. 2014. Multilingual models for compositional distributed semantics. In *ACL*, pages 58–68.

Felix Hill, Roi Reichart, and Anna Korhonen. 2015. SimLex-999: Evaluating semantic models with (genuine) similarity estimation. *Computational Linguistics*, 41(4):665–695.

Adam Kilgarriff. 1997. Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2):135–155.

Alexandre Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing crosslingual distributed representations of words. In *COLING*, pages 1459–1474.

Angeliki Lazaridou, Georgiana Dinu, and Marco Baroni. 2015. Hubness and pollution: Delving into cross-space mapping for zero-shot learning. In *ACL*, pages 270–280.

Ira Leviant and Roi Reichart. 2015. Separated by an un-common language: Towards judgment language informed vector space modeling. *CoRR*, abs/1508.00106.

Omer Levy and Yoav Goldberg. 2014a. Dependency-based word embeddings. In *ACL*, pages 302–308.

Omer Levy and Yoav Goldberg. 2014b. Linguistic regularities in sparse and explicit word representations. In *CoNLL*, pages 171–180.

Omer Levy, Yoav Goldberg, and Ido Dagan. 2015. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the ACL*, 3:211–225.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *ACL*, pages 768–774.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Bilingual word representations with monolingual quality in mind. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 151–159.

André F. T. Martins, Miguel B. Almeida, and Noah A. Smith. 2013. Turning on the turbo: Fast third-order non-projective turbo parsers. In *ACL*, pages 617–622.

Yashar Mehdad, Matteo Negri, and Marcello Federico. 2011. Using bilingual parallel corpora for cross-lingual textual entailment. In *ACL*, pages 1336–1345.

Oren Melamud, David McClosky, Siddharth Patwardhan, and Mohit Bansal. 2016. The role of context types and dimensionality in learning word embeddings. In *NAACL-HLT*, pages 1030–1040.

Rada Mihalcea, Ravi Sinha, and Diana McCarthy. 2010. SemEval-2010 task 2: Cross-lingual lexical substitution. In *SEMEVAL*, pages 9–14.

Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013a. Exploiting similarities among languages for machine translation. *CoRR*, abs/1309.4168.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119.

Bhaskar Mitra, Eric T. Nalisnick, Nick Craswell, and Rich Caruana. 2016. A dual embedding space model for document ranking. *CoRR*, abs/1602.01137.

Joakim Nivre et al. 2015. Universal Dependencies 1.4. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague.

Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Computational Linguistics*, 33(2):161–199.

Yves Peirsman and Sebastian Padó. 2010. Cross-lingual induction of selectional preferences with bilingual vector spaces. In *NAACL*, pages 921–929.

Pushpendre Rastogi, Benjamin Van Durme, and Raman Arora. 2015. Multiview LSA: Representation learning via generalized CCA. In *NAACL-HLT*, pages 556–566.

Hinrich Schütze. 1993. Part-of-speech induction from scratch. In *ACL*, pages 251–258.

Roy Schwartz, Roi Reichart, and Ari Rappoport. 2016. Symmetric patterns and coordinations: Fast and enhanced representations of verbs and adjectives. In *NAACL-HLT*, pages 499–505.

Anders Søgaard, Željko Agić, Héctor Martínez Alonso, Barbara Plank, Bernd Bohnet, and Anders Johannsen. 2015. Inverted indexing for cross-lingual NLP. In *ACL*, pages 1713–1722.

Chen-Tse Tsai and Dan Roth. 2016. Cross-lingual wikification using multilingual embeddings. In *NAACL-HLT*, pages 589–598.

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer. 2015. Evaluation of word vector representations by subspace alignment. In *EMNLP*, pages 2049–2054.

Joseph P. Turian, Lev-Arie Ratinov, and Yoshua Bengio. 2010. Word representations: A simple and general method for semi-supervised learning. In *ACL*, pages 384–394.

Shyam Upadhyay, Manaal Faruqui, Chris Dyer, and Dan Roth. 2016. Cross-lingual models of word embeddings: An empirical comparison. In *ACL*, pages 1661–1670.

Jason Utt and Sebastian Padó. 2014. Crosslingual and multilingual construction of syntax-based vector space models. *Transactions of the ACL*, 2:245–258.

Ivan Vulić and Anna Korhonen. 2016a. Is "universal syntax" universally useful for learning distributed word representations? In *ACL*, pages 518–524.

Ivan Vulić and Anna Korhonen. 2016b. On the role of seed lexicons in learning bilingual word embeddings. In *ACL*, pages 247–257.

Ivan Vulić and Marie-Francine Moens. 2013a. Cross-lingual semantic similarity of words as the similarity of their semantic word responses. In *NAACL-HLT*, pages 106–116.

Ivan Vulić and Marie-Francine Moens. 2013b. A study on bootstrapping bilingual vector spaces from non-parallel data (and nothing else). In *EMNLP*, pages 1613–1624.

Ivan Vulić and Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In *SIGIR*, pages 363–372.

Yogarshi Vyas and Marine Carpuat. 2016. Sparse bilingual word representations for cross-lingual lexical entailment. In *NAACL-HLT*, pages 1187–1197.

Will Y. Zou, Richard Socher, Daniel Cer, and Christopher D. Manning. 2013. Bilingual word embeddings for phrase-based machine translation. In *EMNLP*, pages 1393–1398.