

# Improving Neural Knowledge Base Completion with Cross-Lingual Projections

Patrick Klein and Simone Paolo Ponzetto and Goran Glavaš

Data and Web Science Group

University of Mannheim

B6, 26, DE-68159, Mannheim, Germany

patrick.pat.klein@gmail.com

{simone, goran}@informatik.uni-mannheim.de

## Abstract

In this paper we present a cross-lingual extension of a neural tensor network model for knowledge base completion. We exploit multilingual synsets from BabelNet to translate English triples to other languages and then augment the reference knowledge base with cross-lingual triples. We project monolingual embeddings of different languages to a shared multilingual space and use them for network initialization (i.e., as initial concept embeddings). We then train the network with triples from the cross-lingually augmented knowledge base. Results on WordNet link prediction show that leveraging cross-lingual information yields significant gains over exploiting only monolingual triples.

## 1 Introduction

In the recent years we have witnessed an impressive amount of work on the automatic construction of wide-coverage Knowledge Bases (KBs), ranging from Web-scale machine reading systems like NELL (Carlson et al., 2010) all the way through large-scale ontologies like DBpedia (Bizer et al., 2009), YAGO (Hoffart et al., 2013), and BabelNet (Navigli and Ponzetto, 2012b) as a multi-lingual KB covering a wide range of languages. All KBs, however, are incomplete. Researchers have tried to remedy for the issues of KB incompleteness by constructing knowledge bases of ever increasing coverage directly from the Web (Wu et al., 2012; Gupta et al., 2014; Dong et al., 2014) or by involving community efforts (Bollacker et al., 2008).

Neural models have recently been ubiquitously applied to various NLP tasks, and knowledge base completion (KBC) is no exception (Bordes et al., 2011; Jenatton et al., 2012; Bordes et al., 2013;

Socher et al., 2013; Wang et al., 2014; Yang et al., 2015). These models represent KB concepts and relations as vectors, matrices, and most expressive of them, like that of Socher et al. (2013), as three-dimensional tensors. However, none of these models so far tried to exploit cross-lingual knowledge, i.e., informational and linguistic links between different languages.

We set to fill this gap and propose a cross-lingual extension of the neural tensor network model for knowledge base completion, proposed by Socher et al. (2013) (NTNKBC, henceforth). We develop an approach that grounds entities of the multilingual KB in a shared multilingual embedding space obtained from monolingual word embeddings using the translation matrix model (Mikolov et al., 2013a). We then exploit cross-lingual triples from BabelNet (Navigli and Ponzetto, 2012a), a multilingual knowledge graph as additional information for training the NTNKBC model. Our results show that joining forces across languages and semantics of their corresponding embedding spaces yields significant performance improvements over using monolingual signal only. We believe that a shared multilingual embedding space and cross-lingual knowledge links provide a form of additional regularization for the neural tensor network model and allow for better generalization, consequently yielding significant link prediction improvements.

## 2 Related Work

In recent years a large body of work has focused on knowledge base completion (Yang et al., 2015; Nickel et al., 2016a). *External* KBC approaches use outer knowledge like text corpora (Snow et al., 2012; Aprosio et al., 2013) or other KBs (Wang et al., 2012; Bryl and Bizer, 2014) for acquiring additional knowledge. The text-based external methods typically employ a form of a distant supervi-

sion. They first recognize mentions of pairs of KB entities in text and observe what textual patterns hold between them. They then associate the recognized patterns to particular KB relations and finally search the corpus for other entity pairs mentioned using the same patterns (Snow et al., 2004; Snow et al., ; Mintz et al., 2009; Aprosio et al., 2013). A slight modification is the approach by (West et al., 2014) where lexicalized KB relations are posed as queries to a search engine and results are parsed to find pairs of entities between which the initially queried relation holds. Complementary to this, open information extraction methods (Etzioni et al., 2011; Faruqui and Kumar, 2015) extract large amounts of facts from text that can then be used for extending KBs (Dutta et al., 2014).

Text-centered approaches, however, simply cannot capture knowledge that is rarely made explicit in texts. For example, much of the common-sense knowledge that is obvious to people such as, for instance, that *bananas are yellow* or that *humans breath* are rarely (or never) made explicit in textual corpora. A partial solution to this problem is provided by *internal approaches* that primarily rely on existing information in the KB itself (Bordes et al., 2011; Jenatton et al., 2012; Socher et al., 2013; Nickel et al., 2016b, *inter alia*) to simultaneously learn continuous representations of KB concepts and relations. These models exploit the KB structure as the ground truth for supervision. Obtaining meaningful concept and relation embeddings allows these models to infer additional KB facts from existing ones in an algebraic fashion.

KBs and text are truly synergistic sources of knowledge, as shown by complementary work from Faruqui et al. (2015), who improve the quality of semantic vectors based on lexicon-derived relational information. Internal models for KB completion, however, make no use of cross-lingual links between entities, which are readily available in existing multilingual resources like BabelNet (Navigli and Ponzetto, 2012b). Here, we extend the model of Socher et al. (2013) with cross-lingual links from BabelNet and demonstrate how introducing additional (cross-lingual) knowledge through these links improves the reasoning over the KB in terms of better performance on the link prediction task. Our findings are, in turn, different yet complementary to those found by building cross-lingual embeddings using parallel or comparable data (Upadhyay et al., 2016) or KB-centric multilin-

gual joint approaches to word understanding like, for instance, that of Navigli and Ponzetto (2012b). Assuming that each monolingual embedding space captures a slightly different aspect of a relation between same concepts, by introducing cross-lingual links over a shared embedding space we believe we provide an additional external regularization mechanism for the NTNKBC model.

### 3 Cross-Lingual Information for Knowledge Base Completion

In Figure 1 we highlight the main steps of our cross-lingual extension of the NTNKBC model. We first use BabelNet to translate KB triples used to train the NTNKBC model to other languages. Next we induce the multilingual embedding space by translating monolingual embedding spaces using the linear translation model (Mikolov et al., 2013a). Finally, we build cross-lingual triples and use them as training data for the NTNKBC model.

**Knowledge base translation.** We translate an input monolingual knowledge base  $KB_s$  in the source language  $s$ , e.g., the English WordNet (Fellbaum, 1998), to each target language  $t \in T$  of interest by associating  $KB_s$  concepts and entities with those within a multilingual lexical knowledge resource, e.g., BabelNet synsets (our approach, however, can be used with any multilingual KB providing adequate lexicographic coverage). Multilingual synsets allow us to translate the triples in  $KB_s$  into any of the languages covered by BabelNet. That is, we can translate source language triples  $(e_1^s, r, e_2^s)$  into the corresponding target language triples  $(e_1^t, r, e_2^t)$  for each target language.

**Multilingual embedding space.** We independently train monolingual word embeddings for each of the languages in  $L = \{s\} \cup T$ . Training monolingual word embeddings for each language separately gives us mutually non-associated embedding spaces, which do not necessarily contain similar embeddings for the same concept across languages (e.g., for English word “*cat*” and German word “*Katze*”). This is why we need to project embedding spaces of different languages to a shared multilingual embedding space. To this end, we use the linear mapping model of Mikolov et al. (2013a), where we learn a translation matrix  $M \in \mathbb{R}^{d_t \times d_s}$  (where  $d_s$  is the size of word embeddings of the source and  $d_t$  of the target language) that projects source language embeddings into the embedding

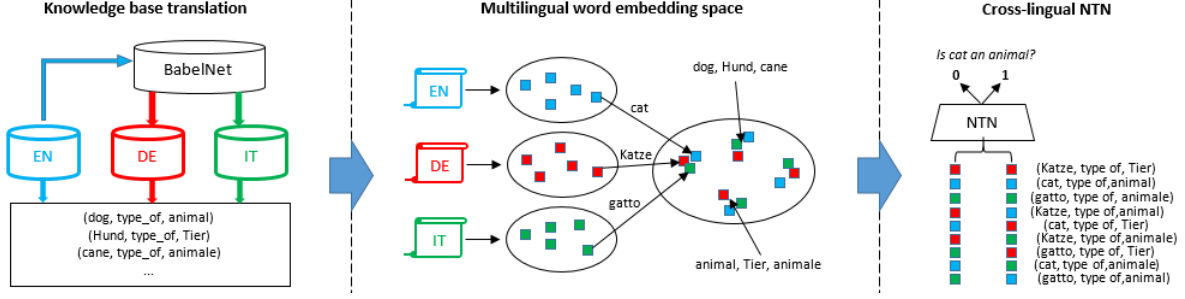


Figure 1: Cross-lingual extension of the NTNKBC model.

space of the target language. Given the training set of word translation pairs of monolingual embeddings  $\{s_i, t_i\}_{i=1}^n$ ,  $M$  is obtained by minimizing the following objective:

$$\sum_{i=1}^n \|Ms_i - t_i\|^2,$$

The obtained matrix  $M$  can then be used to map the embedding of any word from the source language to the embedding space of the target language. To obtain a shared multilingual embedding space we define the embedding space of one of the languages as the target embedding space and project embedding spaces of all other languages to that space. We train one matrix  $M_{t,s}$  for each language  $t \in T$  that we translate  $KB_s$  into, and use it to project the embeddings of  $KB_t$  entities into the same embedding space as that of the source language  $s$ .

**Neural tensor networks for knowledge base completion.** The NTNKBC model of Socher et al. (2013) models KB relations as tensors that bilinearly link KB entities, adding them to the linear associations between entities introduced by earlier models (Bordes et al., 2011). The NTN model assigns the following score to each KB triple  $(e_1, r, e_2)$ :

$$s(e_1, r, e_2) = u_r^T f(e_1^T \mathbf{W}_r^{1:k} e_2 + V_r [e_1] + b_r)$$

where  $\mathbf{W}_r^{1:k} \in \mathcal{R}^{d \times d \times k}$  is the relation-specific tensor for relation  $r$  and  $e_1^T \mathbf{W}_r^{1:k} e_2$  is the bilinear tensor product of entity embeddings  $e_1$  and  $e_2$  that results in a  $k$ -dimensional vector in which each element is computed using a different slice  $W_r^i$  of the tensor  $\mathbf{W}_r^{1:k}$ . Matrix  $V_r \in \mathcal{R}^{k \times 2d}$  linearly links the entities,  $b_r \in \mathcal{R}^k$  is a bias vector, and  $u_r \in \mathcal{R}^k$  is a vector of output layer weights. Relation-specific tensors allow for the multi-perspective modeling of KB relations, with each tensor slice capturing one

aspect of the observed relation. For example, for the relation “*part of*”, one slice might learn that *animals* have *limbs* (from triples like  $(arm, part\ of, person)$ ), whereas another slice could capture that *machines* have *mechanical parts* (from examples like  $(engine, part\ of, car)$ ).

Parameter values, including relation tensors and entity embeddings, are computed by minimizing the cost function  $J(\Omega)$  that couples each correct triple  $F^i = (e_1^i, r^i, e_2^i)$  with corrupt triples  $F_c^i = (e_1^i, r^i, e_c^i)$  in which one entity is replaced with a random KB entity. The correct triples are expected to be scored higher than corrupt triples, which is imposed by forming a standard margin-based objective (i.e., a perfect model will score each correct triple better by at least 1 than any of its corresponding corrupt triples):

$$J(\Omega) = \sum_{i=1}^N \sum_{c=1}^C \max(0, 1 - s(F^i) + s(F_c^i)) + \lambda \|\Omega\|^2$$

where  $\Omega = \{W, V, U, b, E\}$  is the set of all parameters,  $N$  is the size of the training set,  $C$  is the number of corrupt triples for each correct triple, and  $\lambda$  is the regularization coefficient.

**Cross-lingual neural tensor network.** We extend the NTNKBC with multilingual and cross-lingual KB projections. Our hunch is that triples lexicalized in different languages can provide complementary evidence for the existence of a semantic relation between entities (cf. Section 4). Let  $KB_{t_i}$  be the translation of the initial knowledge base  $KB_s$  from the source language  $s$  into the target language  $t_i$ ,  $t_i \in \{t_1, \dots, t_k\}$ . Our new cross-lingual knowledge base (CLKB) then contains:

1. All triples from  $KB_s$ ;
2. All monolingual triples from each of the translated KBs  $KB_{t_i}$ ;

- Cross-lingual triples obtained from monolingual triples by replacing one of the entities with its corresponding entity in another language.

Formally, for each original triple  $(e_1^s, r, e_2^s)$ , CLKB contains  $k$  additional monolingual triples  $(e_1^{t_i}, r, e_2^{t_i})$  and  $2^{\binom{k+1}{2}}$  corresponding cross-lingual triples –  $(e_1^{l_i}, r, e_2^{l_j})$  and  $(e_1^{l_j}, r, e_2^{l_i})$  for each pair of languages  $(l_i, l_j) \in L \times L, i \neq j$ , where  $L = \{s\} \cup T$ . For example, from the English triple *(football player, type of, athlete)* and its corresponding German triple *(Fußballspieler, type of, Sportler)*, we add the following cross-lingual triples *(Fußballspieler, type of, athlete)* and *(football player, type of, Sportler)* to the augmented cross-lingual knowledge base.

Following the NTKNBC approach, we initialize the embeddings of multi-word KB entities by averaging the embeddings of their constituent words (Socher et al., 2013). Finally, we translate the monolingual embeddings of all CLKB entities (obtained from respective monolingual word embeddings) to the shared embedding space and train the NTKNBC model on the CLKB triples.

## 4 Evaluation

In line with previous work (Chen et al., 2013; Socher et al., 2013), we evaluate our approach on the *link prediction* task, namely the binary classification task of predicting the correctness of a KB triple  $(e_1, r, e_2)$ , given entities  $e_1$  and  $e_2$  and a semantic relation  $r$ .

### 4.1 Experimental Setting

**Dataset.** We perform the evaluation on WordNet (Fellbaum, 1998) (i.e., *WN11* dataset), following the same evaluation setting, i.e., the same train, development, and test split as in the evaluation of the original NTKNBC model (Socher et al., 2013). We translate the *WN11* dataset to German (*WN11DE*) and Italian (*WN11IT*) via multilingual BabelNet synsets. Because not all *WN11* synsets have German and Italian counterparts in BabelNet,<sup>1</sup> *WN11DE* and *WN11IT* are somewhat smaller than *WN11*. The sizes of train, development, and test portions (in terms of number of correct triples) are given in Table 1 for each of the three monolingual *WN11* datasets.

<sup>1</sup>Cf. Navigli and Ponzetto (2012a) reporting a synset coverage of almost 70% for German and Italian (Table 6).

WN	#ent	train	dev	test
<i>WN11</i>	38,696	112,581	2,609	10,544
<i>WN11DE</i>	33,353	91,711	2,295	9,213
<i>WN11IT</i>	33,397	91,933	2,295	9,236

Table 1: Sizes of *WN11* datasets.

Mapping	P@1	P@5
DE→EN	36%	53%
IT→EN	40%	58%

Table 2: Embedding translation performance.

**Word embeddings.** We used the WaCky corpora (Baroni et al., 2009) – UkWaC, DeWaC, and ItWaC – to respectively train English, German, and Italian embeddings. We built the 100-dimensional embeddings using the CBOW algorithm (Mikolov et al., 2013b). We then mapped the German and Italian embeddings into the English embedding space by (1) translating 1100 most frequent English words (1000 pairs for training and 100 for testing) to German and Italian using Google translate and (2) training the respective German-to-English and Italian-to-English translation matrices. The quality of the obtained translations, measured in terms of P@1 and P@5 (i.e., percentage of cases in which the translation pair was retrieved as the most similar or among the five most similar words from the other language), is shown Table 2. The performance levels we obtain are comparable to translation performances reported in the original work (Mikolov et al., 2013a).

**Model configuration.** The augmented CLKB contains a total of 846K triples (296K monolingual and 550K cross-lingual). Following (Socher et al., 2013), we set the number of tensor slices to  $k = 4$  and the corruption rate (i.e., number of corrupt triples per each correct triple) to  $C = 10$ . We also optimize the NTKNBC’s parameters with the minibatched L-BFGS algorithm, with minibatches of size  $N = 20.000$  triples. We use the development portion of the *WN11* dataset to optimize the model hyperparameters – the prediction thresholds for each of the 11 types of relations. Finally, we evaluate different model variants on the test portion of *WN11*.

**Models in evaluation.** Different model variants that we evaluate mutually differ only in terms of

Model	Acc.	Prec.	Rec.	$F_1$
MONO-EN	85.82	87.07	84.12	85.57
MONO-DE	83.37	86.06	81.26	83.59
MONO-IT	84.80	86.96	83.38	85.13
ML-NTN	84.60	85.95	82.73	84.30
CL-NTN	<b>87.86</b>	<b>87.94</b>	<b>87.76</b>	<b>87.85</b>

Table 3: Performance (%) on link prediction.

the subset of CLKB triples used for training. The final evaluation of the model is always performed on the triples from the test portion of the original (i.e., English) WN11 dataset. We evaluate:

1. Three monolingual models – MONO-EN (direct reimplement of the original NTNKBC model), MONO-DE, and MONO-IT – trained respectively using only monolingual English, German, and Italian triples;
2. The multilingual model (ML-NTN), trained using the union of the three sets of monolingual triples;
3. The cross-lingual model (CL-NTN) in which we use all cross-lingual triples in addition to all monolingual triples.

## 4.2 Results and Discussion

The link prediction performance for all above-mentioned models, measured on the test portion of the original WN11 dataset (containing English triples) is shown in Table 3.

MONO-EN achieves accuracy of 85.8%, which is very close to the 86.2% accuracy reported by Socher et al. (2013). The monolingual English model MONO-EN significantly ( $p < 0.01$ )<sup>2</sup> outperforms the other two monolingual models. We credit this performance gap to the significantly larger training set (38.7K entities and 112.5K triples vs. 33.4K entities and 92K triples for both German and Italian). The Italian monolingual model (MONO-IT) outperforms the German monolingual model (MONO-DE) despite comparable training set sizes, which we credit to the lower quality of the DE→EN translation matrix in comparison with the IT→EN translation matrix (see Table 2).

The multilingual model outperforms only one of the three monolingual models. This is not so surprising (although it might seem so at first glance) if

<sup>2</sup>All performance differences were tested for significance using the non-parametric stratified shuffling test (Yeh, 2000).

we consider that ML-NTN merely combines three disjoint KBs which share semantic information only through shared embedding space and relation tensors. Without the direct, cross-lingual links between entities of different monolingual KBs, these signals seem to be insufficient to compensate for a much larger number of parameters (three times larger number of entities) that the ML-NTN model has to learn compared to monolingual models.

The cross-lingual model (CL-NTN), on the other hand, significantly outperforms all monolingual models. We believe that this is because by adding cross-lingual triples we introduce additional regularization to the model – although cross-lingual triples describe the same facts as monolingual triples (i.e., same relations between same entities) the facts get represented slightly differently due to imperfect embedding translation and inherent language differences. We believe that this effect is similar to adding noise when training denoising autoencoders (Vincent et al., 2008), in order to obtain more robust entity representations. We believe that the addition of German and Italian monolingual triples has the same regularizing effect as the addition of cross-lingual triples, but their number is significantly smaller (184K compared to 550K cross-lingual triples) and alone they do not compensate for increased model complexity (i.e., three times larger number of entity vectors to be learned).

## 5 Conclusion

We presented a cross-lingual extension of the NTNKBC model of Socher et al. (2013) that leverages a multilingual knowledge graph and multilingual embedding space. Our results indicate that using cross-lingual links between entity lexicalizations in different languages yields better NTNKBC model. That is, our experiments imply that the cross-lingual signal enabled through the multilingual KB and shared multilingual embedding space provides improved regularization for the neural KBC model. We intend to investigate whether such cross-lingual regularization can yield similar improvements for other neural KBC models and whether it can be combined with other types of regularization, such as that based on augmenting KB paths (Gua et al., 2015). We will also evaluate the cross-lingually extended KB-embedding models on other high-level tasks such as error detection and KB consistency checking.

## References

- Alessio Palmero Aprosio, Claudio Giuliano, and Alberto Lavelli. 2013. Extending the coverage of DBpedia properties using distant supervision over Wikipedia. In *Proceedings of the 2013 Workshop on Natural Language Processing and DBpedia*, pages 20–31, Trento, Italy.
- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3):209–226.
- Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. DBpedia – A crystallization point for the web of data. *Journal of Web Semantics*, 7(3):154–165.
- Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 1247–1250, vancouver, British Columbia, 2008.
- Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Proceedings of the 2011 AAAI Conference on Artificial Intelligence*, pages 301–306, San Francisco, California, USA.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 2013 Annual Conference on Neural Information Processing Systems (NIPS)*, pages 2787–2795, Lake Tahoe, Nevada, USA.
- Volha Bryl and Christian Bizer. 2014. Learning conflict resolution strategies for cross-language Wikipedia data fusion. In *Proceedings of the 2014 World Wide Web Conference (WWW)*, pages 1129–1134, Seoul, Korea.
- Andres Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Jr. Hruschka, and Tom M. Mitchell. 2010. Toward an architecture for never-ending language learning. In *Proceedings of the 2010 AAAI Conference on Artificial Intelligence*, pages 1306–1313, Atlanta, Georgia, USA.
- Danqi Chen, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. 2013. Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. In *Proceedings of the Workshop Track of the International Conference on Learning Representations (ICLR)*, page N/A, Scottsdale, Arizona, USA.
- Xin Dong, Evgeniy Gabilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmman, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 2014 ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 601–610, New York City, New York, USA.
- Arnab Dutta, Christian Meilicke, and Simone Paolo Ponzetto. 2014. A probabilistic approach for integrating heterogeneous knowledge sources. In *Proceedings of the 2014 European Semantic Web Conference (ESWC)*, pages 286–301, Anisara/Hersonissou, Crete, Greece.
- Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam. 2011. Open information extraction: The second generation. In *Proceedings of the 2011 International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3–10, Barcelona, Spain.
- Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (NAACL HLT)*, pages 1351–1356, Denver, Colorado, USA.
- Manaal Faruqui, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard H. Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (NAACL HLT)*, pages 1606–1615, Denver, Colorado, USA.
- Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.
- Rahul Gupta, Alon Halevy, Xuezhi Wang, Steven Whang, and Fei Wu. 2014. Biperpedia: An ontology for search applications. In *Proceedings of the 2014 International Conference on Very Large Data Bases (VLDB)*, pages 505–516, Hangzhou, China.
- Kelvin Guu, John Miller, and Percy Liang. 2015. Traversing knowledge graphs in vector space. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 318–327, Lisbon, Portugal.
- Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, pages 28–61.
- Rodolphe Jenatton, Nicolas L. Roux, Antoine Bordes, and Guillaume R. Obozinski. 2012. A latent factor model for highly multi-relational data. In *Proceedings of the 2012 Annual Conference on Neural Information Processing Systems (NIPS)*, pages 3167–3175, Lake Tahoe, Nevada, USA.

- T. 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, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 2013 Annual Conference on Neural Information Processing Systems (NIPS)*, pages 3111–3119, Lake Tahoe, Nevada, USA.
- Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP)*, pages 1003–1011, Suntec, Singapore.
- Roberto Navigli and Simone P. Ponzetto. 2012a. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.
- Roberto Navigli and Simone Paolo Ponzetto. 2012b. Joining forces pays off: Multilingual joint Word Sense Disambiguation. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1399–1410, Jeju Island, Korea.
- Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016a. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33.
- Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. 2016b. Holographic embeddings of knowledge graphs. In *Proceedings of the 2016 AAAI Conference on Artificial Intelligence*, pages 1955–1961, Phoenix, Arizona, USA.
- Rion Snow, Dan Jurafsky, and Andrew Ng. Semantic taxonomy induction from heterogeneous evidence. In *Proceedings of the Joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (COLING-ACL)*, pages 801–808, Sydney, Australia.
- Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2004. Learning syntactic patterns for automatic hypernym discovery. In *Proceedings of the 2004 Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1297–1304, Vancouver, British Columbia, Canada.
- Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Proceedings of the 2013 Annual Conference on Neural Information Processing Systems (NIPS)*, pages 926–934, Lake Tahoe, Nevada, USA.
- Shyam Upadhyay, Manaal Faruqui, Chris Dyer, and Dan Roth. 2016. Cross-lingual models of word embeddings: An empirical comparison. In *Proceedings of the 2016 Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 1661–1670, Berlin, Germany.
- Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 2008 International Conference on Machine Learning (ICML)*, pages 1096–1103, Helsinki, Finland.
- Zhichun Wang, Juanzi Li, Zhigang Wang, and Jie Tang. 2012. Cross-lingual knowledge linking across wiki knowledge bases. In *Proceedings of the 2012 World Wide Web Conference (WWW)*, pages 459–468, Lyon, France.
- Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the 2014 AAAI Conference on Artificial Intelligence*, pages 1112–1119, Québec, Canada.
- Robert West, Evgeniy Gabrilovich, Kevin Murphy, Shaohua Sun, Rahul Gupta, and Dekang Lin. 2014. Knowledge base completion via search-based question answering. In *Proceedings of the 2014 World Wide Web Conference (WWW)*, pages 515–526, Seoul, Korea.
- Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 International Conference on Management of Data (SIGMOD)*, pages 481–492, Scottsdale, Arizona, USA.
- Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the 2015 International Conference on Learning Representations (ICLR)*, page N/A, San Diego, California, USA.
- Alexander Yeh. 2000. More accurate tests for the statistical significance of result differences. In *Proceedings of the 2000 International Conference on Computational Linguistics (COLING)*, pages 947–953, Saarbrücken, Germany.